THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME VIII (1941)

GENERAL EDITOR
W. B. MANN

Price to non-Fellows 25*s. net*

Reproduced and Printed Photo-Litho Offset for
WM. DAWSON & SONS LTD.
16, West Street, Farnham, Surrey
1960

With the permission of
THE PHYSICAL SOCIETY

THE GENERAL PHYSICAL CONSTANTS

By RAYMOND T. BIRGE

Reports on Progress in Physics, vol. VIII (1941)

CORRIGENDA

Page 90, last line. Insert "most" before "dependent".

,, 103, line following equation (3). Insert "limiting" before "value".

,, 107, line 10 from bottom. The sentence is clarified by the substitution of "the gamma value" for "one gamma".

,, 112, line 4. J_{15} should be J_{18}.

,, 112, line 5. • should be †.

,, 112, line 20. † should be •.

,, 120, last line of table 2. "Goedieke" should be "Goedicke".

,, 121, first line of first footnote. "1932 b" should be "1932 c".

,, 122, footnote. Insert the symbol •.

,, 123, line 15 from bottom. "Birge, 1940" should be "Birge, 1940 b".

,, 125, first line of last paragraph. "$h/e' \pm \ldots\ldots$" should be "$h/e' = \ldots\ldots$".

,, 126, Table *a*. In the unit of G, "cm." should be "cm^3".

,, 130, Radiation density constant. The numerator "$4\pi N_0 \ldots\ldots$" should be "$4\pi N^3_0 \ldots\ldots$".

,, 133, Manian, Urey and Bleakney. "1934" should be "1935".

CONTENTS

	PAGE
Preface	iv
The teaching of theoretical physics in universities. By M. Born	1
The theory of the specific heat of solids. By M. Blackman	11
High-speed centrifuging. By J. W. Beams	31
Applications of spectroscopy to combustion. By A. G. Gaydon	50
Friction between solid bodies. By Robert Schnurmann	71
The general physical constants. By Raymond T. Birge	90
Recent television developments. By V. K. Zworykin and R. E. Shelby	135
Instrumental technique in astrophysics. By A. Hunter	186
Dispersion in the far infra-red. By Lotte Kellner	200
The M.I.T. wavelength project. By George R. Harrison	212
Molecular electronic spectra, dispersion and polarization. By Robert S. Mulliken and Carol A. Rieke	231
Symmetry properties of nuclear levels. By Eugene P. Wigner and Eugene Feenberg	274
Photometry. By H. Buckley	318
Discharge phenomena in gases. By R. Winstanley Lunt, A. von Engel and J. M. Meek	338
Author index	368

PREFACE

The task of soliciting *Reports* in war-time is comparable with that of the biblical maker of bricks after a lean wheat year. In the circumstances, therefore, I must acknowledge with gratitude the unstinted support and help which have been accorded by colleagues and friends from both this and the other side of the Atlantic. Several American physicists not only contributed articles but also asked that the honoraria attached thereto should be donated to any suitable and worthy fund in this country. Such disinterested and generous actions demonstrate, with no uncertainty, the ties of friendship and of sympathy linking physicists in the English-speaking world, and are happy auguries of happier times to come.

Further to the introduction, last year, of the Harvard system of reference, another slight innovation has now been made. In a volume of this kind there is something to be said for the numeralogical method of reference in that an author's name can the more readily be located in the text. This is not so easily done with the Harvard system, but this difficulty has been eradicated in the present volume by the introduction of an author index. This index is not complete in that it contains only the names of those authors listed in the various bibliographies. With this volume is also issued the list of references for the *Report* in volume VII by F. Eirich on the *Viscosity and Nature of Substances of High Molecular Weight in Solution*. This list is printed on a single slip which can easily be inserted between pages 354 and 355 of volume VII.

The abrupt change in the quality of the paper toward the end of this volume, which is much to be regretted, is a consequence of war conditions.

On account of the current delays in communication between this country and America, most of the articles by American authors were sent to press without a corrected proof having been received from the respective authors. In order to reduce errors to a minimum these articles were also variously read, or glanced through, by a number of colleagues. For most invaluable help of this nature I have particularly to thank Dr. C. A. Coulson, Dr. M. Blackman, Dr. W. D. Wright, and Dr. R. W. B. Pearse. For equally valuable help I am also especially indebted to Mr. J. H. Awbery and Dr. W. Jevons.

The author index was compiled by my wife, but appropriate and adequate appreciation for this assistance has already, and elsewhere, been expressed.

Imperial College,
London, S.W. 7,
April 1942.

W. B. M.

AUTHOR INDEX

Adams, 190, 191.
Albertson, 228, 229.
Allsopp, 264.
Alvarez, 295.
Amontons, 72, 73, 81, 82.
Anderson, 94–101.
Andersson, 40.
Archibald, 32, 33.
Armstrong, 151.
Arndt, 324.
Arnot, 355.
Astbury, 106.
Aston, 35, 36, 311, 349.
Atkinson, J. R., 331.
Atkinson, R. d'E., 192.
Avramenko, 58.
Awbery, 65.

Bacher, 301, 303.
Badami, 57.
Bailey, C. R., 61.
Bailey, H. H., 125, 126.
Bailey, V. A., 347.
Baker, 187, 189.
Baldwin, 167, 172, 173.
Bär, 350.
Barbrow, 319, 322, 329, 331.
Bardeen, 305.
Barkas, 279, 280, 310, 311.
Barnard, 330, 331.
Barnes, B. T., 325, 328.
Barnes, C., 65.
Barnes, R. B., 200–204.
Barnett, 320.
Barratt, 57.
Bartlett, G., 164.
Bartlett, J. H., 277, 301.
Bates, J. R., 57.
Bates, S. J., 114.
Batuecas, 104, 105, 110.
Bauer, 41, 45.
Bawden, 37.
Bawn, 62.
Baxt, 60.
Baxter, 104, 109.
Bayliss, 235.
Beams, H. W., 37, 47.
Beams, J. W., 36–48, 362.
Bearden, 117–126.
Beattie, 110.
Beckmann, 46.
Beeck, 88.
Bedford, 145–148, 172–176.
Beer, 335.
Beers, 141.
Behn, 12.
Bell, 53, 56.
Bender, 22, 28.
Benford, 328.
Berkey, 362.
Bethe, 20, 277–308.
Bhabha, 295.
Bikerman, 72.
Birge, 64, 90–125.
Birkinshaw, 137, 150, 153, 176.
Bitter, 214, 227.
Björnstahl.*
Blackman, 11–28, 204, 207–209.
Blankenstein, 109.
Bleakney, 107, 108.
Bloch, 280, 295.
Blok, 79.
Blumlein, 150, 153, 173, 176.
Bohr, 308.
Bollmann, 118, 125.
Bond, 116, 123–125.
Bone, 56, 60.
Bonhoeffer, 50, 51, 58.
Booker, 259.
Borelius, 77.
Born, 11–30, 200–204.
Boutry, 329.
Bouvard, 114.
Bowden, 72–88.
Bowen, 188.
Bracey, 186.
Bradbury, 347.
Brattain, 203, 204, 206.
de Bray, 92, 97, 99, 100.
Breit, 275–314.
Brenner, 361.
Brewer, 352.
Briegleb, 267.
Brinkman, 362.
Broch, 27.
Brode, 345.
Brodhun, 319.
Brody.*
Brohult, 40, 45, 46.
Brose, 349.
Brown, C. O., 150, 153, 173, 176.
Brown, G. H., 153, 180.
Brown, H., 307.
Brown, S. C., 280.
Brown, W. G., 256, 260.
Brown, W. J., 155.
Brown, W. L., 354.
Bruce, 364.
Buchmüller, 331.
Buckley, 318–335.
Burch, 50, 55.
Burgess, 319.
Butt, 39.

Caldwell, F. R., 319, 322.
Caldwell, P. S., 164.
Campan, 353.
Campbell, 331.
Carroll, 307.
Cartwright, 186, 201, 209, 210.
Casedo, 110.
Casimir, 111, 276.
Cassen, 277.
van Cauwenberghe, 362.
Chaffee, 120.
Chako, 234.
Chang, 102.
Chapman, 35.
Child, 234.
Chiles, 38, 45, 48.
Chiplonkar, 354.
Christie, 190.
Christy, 123, 125.
Chu, 120.
Chukri, 37.
Clark, C. W. 20, 23.
Clark, G. L., 56.
Clews, 123.
Cloud, 361.
Clusius, 20, 21, 23.
Collie, 44.
Collier, 334.
Compton, 349, 352, 353.
Condon, 274–277, 303, 304.
Coolidge, 237, 244, 268.
Cooper, 299.
Coulomb, 72.
Coulson, 237–246.
Cowan, 354.
Cragoe, 104.
Crawford, 334.
Creutz, 280.
Critchfield, 105, 297.
Crittenden, 324.
Curtis, H. D., 197.
Curtis, L. L., 105, 106.
Curtis, R. W., 105.
Czerny, 200–204, 208, 209.

Darrow, 365 (Frey).
David, 61, 65.
Davis, C. O., 65.
Davis, D. N., 228, 229.
Davis, N. E., 150, 153, 173, 176.
Davis, T., 39.
Dawsey, 59.
Debye, 11–29.
Delsasso, 280, 281.
Denbigh, 264, 266.
Dennison, 313–315.
Derjaguin, 78, 88.
DeVore.*
Dillon, 362.
Dirac, 281, 308.
Dockerty, 21.
Dole, 108.
Dorsey, 98.
Doubleday, 73.
Downey, 54, 57.
Dresler, 324, 326, 327.
Drinkwater, 120.
Drucker, 47.
Druyvesteyn, 338, 345, 350–354.
Dudding, 327.

Duffendack, 355.
DuMond, 114–126.
Dunnington, 116–123, 361.
Durand, 19, 20.
Dwyer, 67.
Dyer, 172.
Dziobek, 327, 335.

Easthope, 246.
Eddy, 139.
Edmondson, 101.
Egerton, 59.
Ehrenfest, 314.
Einstein, 11, 12.
Eisenbud, 275–277, 283, 296.
von Elbe, 54, 67.
Elford, 38, 45.
Elsasser, 277, 301.
Emeléus, H. J., 53, 54, 57.
Emeléus, K. G., 330–355.
van den Ende, 20, 21, 23.
von Engel, 344, 349, 354, 355.
Engstrom, 150, 154.
Epstein, 152.
Ernst, 88, 349.
Errera, 209.
Escribano, 104.
Espenschied, 142.
Estermann, 73.
Estey, 327.
Eucken, 17–24.
Evans, 259.

Fabry, 327.
Farkas, 52, 276.
Farnsworth, 136, 149, 160.
Faxen, 31.
F., D. G., 150, 166, 170, 179, 180.
Feenberg, 276–314.
Ferenez, 325.
Fermi, 296–298.
Finch, 57.
Fine, 25.
Fink, 137, 145, 159, 172.
Fischer, 354, 355, 356.
Flory, 159.
Flügge, 107.
Fock, 203, 207.
Fogg, 87.
Förster, 253.
Försterling, 16, 17.
Foster, 145, 156.
Forsythe, 325, 327, 328.
Fowler, A., 54, 55, 57.
Fowler, R. H. 12.
Fowler, W. A., 281.
Francis, 338, 353, 354.
Franck, H. H., 60.
Franck, J., 358, 363.
Fredenhall, 145, 147, 172, 173.
Freeth, 331.
Frenkel, 78.
Frey, 349.
Friauf, 65.
von Friesen, 123, 125.
Frohlich, 295.
Frömel, 235.
Frost, 58, 59.
Fuchs, 22, 26, 28, 307.
Fucks, 354.

Gamow, 297–299.
Garman, 37.
Garfitt, 362.
Garner, 61, 62.
Gaydon, 51–68.
Gehrts, 353.
Geib, 54.
Geiger, 356.
Gerjuoy, 296.
Gerlach, 120.
Gerö, 64.
Gershun, 335.
Gerthsen, 354.
Gibson, G. E., 17.
Gibson, K. S., 323, 328, 332.
Gillam, 259,
Gillod, 329.
Ginnings, 111, 112.
Girard, 37.
Givens, 88.
Goddard,[2]
Goedicke, 120.
Goens, 20.
Goeppert-Mayer, 300.
Goldhaber, 280.
Goldman,[9]
Goldmark, 150, 166, 170, 172.
Goldsmith 157, 177.
Gordon, 65.
Gray, 148, 173.
Green, A. E. S., 125, 126.
Green, E., 150, 153, 173, 176.
Greene, 107.
Gregory, 57.
Grenat, 53.
Grieneisen, 257.
Griffiths, 65.
Grönblom, 298, 299.
Grüneisen, 20.
Guggenheimer, 301.
Guichard, 109.
Guild, 334.
Guillaume, 102.
Güntherschulze, 350, 351, 353, 355.
Gutton, 94.
Guy, 151.

Haber, 50, 51.
Hafstad, 313, 324.
Hahn, 107.
Hall, D. A., 62.
Hall, J. S., 192.
Hanson, 137–145.
Hardy, T. C., 294.
Hardy, W. B., 73.
Harkins, 311.
Harper, 329.
Harris, N. I., 354.
Harris, W. A., 344.
Harrison, G. R., 214, 215, 227, 228.
Harrison, H. T., 329.
Harteck, 21, 23, 54.
Hartley, 61, 63.
Hartshorn, 106.
Harvey, E. N., 37, 47.
Harvey, F. E., 62.
Haseltine, 358.
Hausser, 259.
Haxby, 280.
Haykin, 79, 83.
Hecht, 335.
Hedfield, 51.
Heisenberg, 275, 277, 291, 301.
Heitler, 295.
Hendrix, 189.
Herne, 56.
Henning, 101, 109, 110.
Henri, 234, 235, 251, 259.
Henriot, 37, 38.
Hercus, 112.
Herold, 145, 155.
Herzberg, 52–66, 270, 271.
Herzfeld, 241.
Heyl, 101.
Hickok, 159.
Higgs, 347.
Hiltner, 193.
von Hippel. 358, 363.
de la Hire, 72.
Hirschfelder, 246.
Hitchcock, 228.
Hoffer, 259.
Hoffman, 319.
Hohls, 203.
Hollywood, 153.
Holm, 87, 358.
Holmes, F. T., 43.
Holmes, R. S., 154.
Holst, 358.
Holstein, 300.
Hönigschmid, 109.
Honnefelder, 18.
Hopf, 18, 19.
Hopper, 119.
Hori, 53.
Horvay, 307.
Hosford, 228.
Houston, 120.
Houtgast, 193.
Hückel, E., 254–261.
Hückel, W., 264.
Hughes, 77, 78, 87.
Huguenard, 37, 38.
Hulme, 198.
Humason, 191.
Humphreys, 36, 42.
Hund, 50, 275, 282, 284, 306.
Hunter, 192.
Hunwicks, 87.
Hüttel, 97.
Hyde, 335.

Iams, 149, 161–168.
Inglis, 294, 307, 313.
Ives, 163, 319–332.

Jack, 51.
Jacobs, 77.
Jaeger, 101, 109–112.
Jaggi, 325.
Jainski, 325.
James, 237, 244, 268.
Janes, 159.
Janossy, 300.
Jenkins, F. A., 104.
Jenkins, H. G., 338.
Jenkins, K. G., 353, 354.
Jensen, 130, 143, 161.
Jevons, 50, 63, 66.
Johnson, 61, 62.
Johnsrud, 163.
Johnston, 65.
Joly, 35.
Jones, G. W., 65.
Jones, H. S., 186.
Jones, R. N., 262.

Kahler, 39.
Kaidanovsky, 83.
Kallman, 152.
Kallmann, 353.
Kapzov, 358.
v. Kármán, 11–25.
Karolus, 96.
Kassel, 65.
Keegan, 332.
Keesom, 20, 21, 23.
Kell, 172, 173.
Keller, F., 353.
Keller, J. M., 123, 125.
Kellermann, 11, 26.
Kellogg, 276, 295.
Kemmer, 295.
Kenty, 354.
Killian, 345.
King, 37, 47.
Kingsbury, 332.
Kinsler, 120.
Kip, 363.
Kirchner, 116, 119, 120, 123.
Kirkby, 338–347.
Kirkpatrick, 123, 125.
Kirschstein, 87.
Kitagawa, 52.
Klemenc, 60.
Knipp, 281.
Kohlrausch, 257.
Kok, 20.
Kondratjew, V., 51–59.
Kondratjewa, E., 56.
Kondratjewa, H., 56.
König, A., 335.
König, H., 328–333.
Konopinski, 282, 299.
Korff, 66.
Korsching, 294.
Korth, 202, 204.
Kraemer, 32, 34.
Kratzer, 53.
Kroeger, 307.
Kroger, 151, 154.
Kuhn, 259.
Kusch, 277, 294.
Kurzke, 354.

Laby, 112, 119.
Lambert, 335.
Lamm, 31, 40, 46.
Landis, 46.
Landshoff, 303.
Lange, 21.
Langmuir, 353, 354.
Lansing, 32, 34.
Larner, 135.
Lauer, 56.
Lauffer, 45.
Lauritsen, 281.
Lavin, 52.
Law, 158, 178.
Lazarew, 88.
Leavey, 77, 78, 87.
Leben, 83, 85, 88.
Lechner, 18, 19.
Lennard-Jones, 253, 256.
Levin, 45.
Lewis, B., 54, 65, 67.
Lewis, G. N., 17, 259, 262.
Leyda, 38.
Lih, 61.
Lindemann, 13, 27, 35, 36.
Lindenblad, 153.
Lindsley, 228.
Link, 38, 41.
Lipkin, 259, 262.
Lissovsky, 79, 83.
Little, 327.
Loeb, 345–363.
Lohr, 137.
Lommell, 335.
Loughren, 148, 172, 173.
Lowry, 264.
Luckeish, 164, 166.
Lunt, 339–347.
Lyddane, 250.
Lyot, 195, 196.

McBain, 37, 38, 41.
McDermott, 329.
McFarlane, 46.
MacHattie, 43, 44.
Machin, 327.
McIlwain, 150, 159, 160.
MacInnes, 47.
McIntosh, 37.
Mack, 228.
McMath, 197, 198.
McMurry, 235–271.
McNally, 214, 228.
Macnamara, 137, 150, 153, 176.
Magel, 259, 262.
Majorana, 281, 301.
Maloff, 158, 172, 178, 179.
Manian, 107, 108.
Margenau, 293, 294, 301, 307.
Martin, 192.
Masket, 42.
Mason, 30.
Massey, 353, 355, 358.
Mattauch, 107.
Mayo, 123, 125.
Meek, C. A., 339–347.
Meek, J. M., 358–364.
Mentzel, 201.
Merchant, 73, 88.
Mercier, 98, 99.
Mertz, 148, 150, 173.
Meyer, 329.
Michelson, 93, 99, 100.
Miller, 118.
Milligan, 355.
Millman, 277, 294.
Mills, 333.
Minkowski, 188.
Minnaert, 193.
Mittelstaedt, 96, 97, 99.
Moles, 104, 105.
Molnar, 214, 228.
Moon, C., 106.
Moon, P., 333, 335.
Morgan, F., 85.
Morgan, J. D., 362.
Morris, 137–151.
Morse, 349, 352.
Morton, 137–178.
Mott, 353, 358.
Motz, 305.
Mountjoy, 155.
Mrowka, 242–246.
Mulders, 193.
Mulliken, 35–53, 232–271.
Murphey, 107.
Murrill, 110.
Muskat, 85.

Nelson, 21, 299.
Nergard, 153.
Nernst, 13, 27.
Newitt, 60.
Newton, 191.
Nichols, 46.
Nord, 362.
Nordheim, G., 271.
Nordheim, L., 297.
Nordsieck, 276.

Ohlin, 125.
Oldenberg, 58, 59, 66, 67, 235.
Oliphant, 354.
O'Neal, 280.
Oosterhuis, 358.
Oppenheimer, 299, 314, 315.
Osborne, 111, 112.
O'Shaughnessy, 262.
Outridge, 56, 60.

Paddock, 235, 259.
Paetow, 354, 362.
Painter, 157, 158, 178.
Palmer, 327, 331.

Panofsky, 125, 126.
Papenhuyzen, 83.
Parodi, 204.
Paterson, 327.
Pauling, 257–261, 267.
Payne-Gaposchkin, 188.
Pease, 93, 99.
Pearse, 51, 53.
Pearson, F., 93, 99.
Pearson, T. G., 340.
Pedersen, 32–47.
Penning, 338, 352, 354, 357.
Perlman, 125.
Perrin, 327.
Perrott, 65.
Perry, 120.
Pettit, 198.
Phillips, C. E. S., 83.
Phillips, M., 294, 296, 303–307.
Philpot, 40.
Pickels, 37–39, 41, 45.
Pickett, 234, 235, 251, 259.
Pidgeon, 59.
Pike, 354.
Pirie, 38.
Poch, 152.
Poincaré, 100.
Pollard, 62.
Pollock, 54.
Pool, 35.
Posin, 363.
Prandtl, 78.
Present, 275, 276, 302, 310.
Preston, 325, 326, 329, 330.
Pretty, 52.
Price, 237, 257.
Primakoff, 300.
Protzman, 137, 140.
Purcell, 54, 57.

Rabi, 276, 294, 295.
Rabinowitch, 235.
Raether, 360, 361, 362.
Ramsey, A. S., 93.
Ramsey, N. F., 276, 295.
Rankin, J. A., 145, 156.
Rankin, J. S., 78.
Rarita, 277, 294–296.
Rassweiler, 56, 57.
Rayleigh, 83, 85, 93.
Reeb, 327.
Reed, 85.
Rees, 235.
Reichardt, H., 58, 60.
Reichrudel, 354, 355.
Ribaud, 320.
Rice, F. O., 59.
Rice, O. K., 67.
Richardson, 120.
Rideal, 72.
Ridler, 84.
Rieke, C. A., 256, 260.
Rieke, F. F., 66, 67, 235.
Riesenfeld, 102.
Riley, 54, 57.
Rimmer, 57.
Roach, 250.
Roberson, 228.
Robinson, C. F., 120.
Robinson, D. M., 130, 157.
Robinson, H. R., 119, 123, 125.
Rodebush, 262.
Roebuck, 110.
Roeser, 319, 322.
Roffey, 62.
Rogers, 250.
Rogowski, 354, 355, 358–363.
Rollefson, 125.
Rompe, 352, 354, 356, 357.
Rosa, 98.
Rose, A., 149, 161, 168.
Rose, M. F., 294.
Rosen, B., 353.
Rosen, N., 67, 228.
Rosenthal, 164.
Ross, F. E., 189.
Ross, P. A., 123, 125.
Roux, 327.
Runge, 353.
Russell, 228, 229.

Sachs, 294, 300.
Sackter, 235, 259.
Safford, 361.
Sakata, 295.
Sämmer, 358.
Sandler, 276.
Sarnoff, 150.
Saunders, 62.
Sayers, 355.
Schade, O. H., 141.
Schade, R., 354, 358.
Scheel, 109.
Scheibe, 235, 257.
Scherzer, 352, 353, 356, 357.
Schmid, 55, 64, 68.
Schmidt, 294.
Schnurmann, 73–87.
Schomaker, 257, 260.
Schön, 352, 354, 356, 357.
Schuhmann, 333.
Schuler, 294.
Schumann, 358.
Schwarz, 126.
Schwers, 21.
Schwinger, 277, 294–296, 315.
Seeley, 151.
Seeliger, 354, 355, 357, 363.
Seitz, 203, 204 206.
Selbie, 37,
Severinghaus, 45.
Shane, 120.
Shapiro, 39.
Share, 281.
Sharp, 329.
Shaw, A. E., 120.
Shaw, P. E., 77, 78, 87.
Shelby, 137, 140, 141.
Shenstone, 222, 355.
Shortley, 274, 303, 304.
Shoupp, 280.
Shrider, 325, 328.
Simon, 21, 22, 28.
Skarstrom, 36, 42, 43.
Skellett, 196.
Skilling, 361.
Sklar, 252, 271.
Slater, 274, 302–304, 314.
Slepian, 362.
Sloane, 354.
Slobod, 108.
Smakula, 259, 262.
Smith, A. E., 88.
Smith, C. S., 43, 44.
Smith, E. C. W., 53, 56, 68, 340–346.
Smith, H. A., 56, 329.
Smith, J. E., 151, 154.
Smith, J. P., 148, 174, 176.
Smith, P. T., 343.
Smith, R. A., 355.
Smith, T. A., 141.
Smithells, 50.
Smythe, 107.
Snoddy, 39, 41, 43.
Solomonovich, 79, 83.
Sommer, 38, 41.
Sommerfeld, 20.
Soyten, 31.
Spain, 307.
Sparks, 106.
Spedding, 120.
Spencer, 152.
Spiwak, 354, 355.
Sponer, 64, 271.
Stanley, 45.
Stark *et al.*, 259.
Starkweather, 104.
Stauffacher, 41.
Steele, 63.
Steenbeck, 344–355.
Steensholt, 246.
Stehn, 275, 276.
Steinhardt, 333.
Steinwehr, H., 110–112.
Stephens, 280.
Stern, 73.
Steubing, 349.
Stevens, 78.
Stewart, 62.
Stiles, 334.
Stimson, 111, 112.
Stockmair, 257,
Stott, 102.
Street, 362.
Strieby, 141, 142, 154.
Strigel, 362.
Stuart, 245.
Sugiura, 246.
Suydam, 98.
Svedberg, 32–47.
Swain, 21.
Swartout, 107, 108.
Swindell, 340, 342, 343.

Tabor, 72, 88.
Taketani, 295.

Tamm, 362.
Tate, 343.
Taylor, A. H., 324
Taylor, J., 358.
Taylor, W. G. A., 334.
Teele, 332, 333.
Teller, 270, 271, 297–314.
Thaxton, 275, 276, 296.
Thee, 56.
Thirring, 15, 17.
Thomson, G. P., 353, 358.
Thomson, J. J., 352, 358.
Thompson, H. H., 57.
Thompson, H. W., 60.
Tiltman, 163.
Timoshenko, 335, 355.
Tiselius, 41.
Tolman, 47, 66.
Tolson, 158, 178, 179.
Tomlinson, 72, 78.
Tonks, 348, 354.
Topley, 340.
Toral, 104.
Townsend, 338, 360.
Trevor, 151, 154.
Trump, 361.
Tutte, 237.
Tyndall, A. M., 3[illegible]
Tyndall, E. P. T.,
Tyrell, 307.
Tyson, 155.

Ubbelohde, 53, 54, 59.
Urbanek, 325.
Urey, 36, 57, 59, 107, 108.

Vaidya, 53, 54, 57.
Varney, 358, 363.
Vaughan, 21.
Vigoureux, 106.
Vinal, 114.
da Vinci, 72.
Van Vleck, 68.
Voge, 68.
Volkoff, 277, 302, 303.
Voogd, 332.
van Voorhis, 355.
Voskuyl, 107.

Wagner, 60.
Waidner, 319.
Walche, 354.
Waldmeier, 196.
Walker, M. K., 65.
Walker, O. J., 234,
Wallraff, 362, 363.
Walsh, A. D., 237, 257.
Walsh, J. W. T., 329, 334.
Warmoltz, 354.
Wassermann, 259, 263.
Ward, 80.
Warlow-Davies, 80, 81.
Watson, 51.
Way, 293.
Weaver, K. S., 335.
Weaver, W., 30, 32, 33.
Wechsberg, 60.
Weed, 37, 38.
Wefelmeier, 313, 314.
Wehnelt, 354.
Weis, 141, 154.
Weizel, 352, 354, 356, 357.
Weizsäcker, 300, 303.
Wells, 280.
Went, 332.
Wensel, 116, 123.
Wentz, 142.
Wenzel, 319, 332.
Werth, 24.
Westhaver, 352.
Weston, 55.
Wheeler, H. A., 145–148, 172, 173.
Wheeler, J. A., 243, 313–315.
Whipple, 188.
White, H. E., 227.
White, H. J., 361–363.
White, J. U., 60, 64, 66, 67.
White, M. G., 280.
Whitmore, 335.
Whytlaw-Gray, 109.
Wick, 302.
Wigner, 274–311.
Williams, R. C., 120, 191, 193.
Williams, R. E., 335.
Williams, W. E., 120.
Wilson, E. B., Jr.*
Wilson, G. H., 36.
Winch, 327, 329–332.
Withrow, 56, 57.
Wolf, F., 353.
Wolf, K. L., 241.
Wood, 235.
Woiteneck, 21.
Wollman, 271.
Wyckoff, 41, 42, 45.
Wynn-Williams, 362.

Yamauti, 327, 335.
Yost, 297.
Yukawa, 295.

Zacharias, 276, 295.
Zeidler, 21.
Zeise, 65.
Ziskin, 51, 52, 58.
Zwikker, 21.
Zworykin, 136–178.

* Not referred to in text.

ADDENDUM

P. 29. Dockerty, S. M., 1937. . *Canad. J. Res.* **15**, Sect. A, 59.

Dutch reprint

THE TEACHING OF THEORETICAL PHYSICS IN UNIVERSITIES

By M. BORN, F.R.S.,
The University, Edinburgh

THE splitting of natural philosophy into so many separate sciences is a consequence of the impact of growing knowledge upon the limitations of the individual brain—a very regrettable, but unavoidable, development. Particularly regrettable is the separation of one single science, physics, into two parts which differ not with respect to the object but with respect to the method. This separation would be dangerous indeed, if it were more than a question of personal qualities—one man being better suited to experimental work, another to mathematical interpretations. But as long as they consider themselves as collaborators at the same task no evil consequences will follow. In fact, the development of physics during the last decades has provided many examples of this collaboration, with marvellous results.

The teaching of theoretical physics must be considered from this standpoint of the unity of science. If there is anything crucial in the theoretical side, it is the problem of the philosophical foundation of physics, which, as a matter of course, will be the centre of attention for those who are theoretically minded; it is this point which justifies a separate discussion of the question of how theoretical physics should be taught.

This teaching has a threefold aim, since it is concerned with three different groups of students. The first seeks a general education. If theoretical physics has any educational value at all, it will be in its bearing upon fundamental questions of philosophy. The problem here is how to introduce our subject in the study for an ordinary degree.

The second group consists of those students who need theoretical methods for professional purposes. These are the physicists proper, the physical chemists, astronomers, mineralogists, and so on, who are intending to enter industry, State laboratories, or to become schoolmasters. The teaching problem for these does not involve questions of principle, but the practically difficult problem of devising a syllabus for an honours degree which adequately covers the whole field of classical and modern theory, without neglecting the intimate connection with the experimental side of the science.

The third group consists of the few outstanding students who wish to devote their lives to science itself, in research and higher teaching. The problem here is how to provide these with sufficient lectures and seminars when most of the time of the staff is occupied by " ordinary " teaching.

I speak of this difficulty from my experience at Edinburgh University. For many generations natural philosophy was an integral part of general education at Scottish universities. I occasionally meet elderly gentlemen of every kind of profession—business men, lawyers, ministers, doctors—who, when learning about my position, begin to recollect their own student days, when they attended Tait's lectures on natural philosophy (which were then included in the syllabus of the ordinary degree), and they all praise the clear explanations and the brilliant experiments which produced unforgettable impressions. I then feel rather jealous; for the system has changed, and what I have to teach is the more or less routine material required by students who want to become useful members of a technical society.

Some kind of introduction to modern science ought to be introduced into our general educational system in the universities as well as in the schools. The fact that our civilization is essentially based on science is at present emphasized in many articles, in the daily press as well as in periodicals (e.g., in *Nature*), and it is recommended that scientists should have a greater share in the responsibility of administration and in decisions about the fate of the nation. It may be doubted whether real scientists, men of specific talent and trained for careful observation and consideration, are very useful in a field where activity and versatility of mind and the power of quick decision are most important. But the knowledge and experience of the scientists can be used for the benefit of society by forming committees of experts, a tendency now growing. The full development of it is only possible if the whole nation becomes "science minded"; the characteristic features of scientific research and the general results of natural philosophy should become an integral part of general education. This demand cannot be discussed without touching upon the general problem of the value of science, for the individual as well as for society. Values, however, are a matter of personal experience and inclination, and the following remarks should be understood in this way.

The essence of the question is contained in the eternal tale of Doctor Faustus, symbolizing the soul of man, wrestling for truth and beauty, and of Mephistopheles, the roguish devil, offering him a worldly paradise which ends in hell. We have the choice of teaching science as natural philosophy, framed into a general Faustian education, or as the creator of wealth and power, with the alternative of such wholesale destruction as we are witnessing to-day.

Many current doctrines deny the former aspect. Most of the philosophical schools disregard science more or less. On the other hand, a few groups, positivists, pragmatists, and so forth, consider scientific thought as the only philosophical one. But they are extremely formalistic, and reduce science to a kind of grammar: discoveries are twisted into definitions, new conceptions into conventions, the meaning of scientific thought degraded into language whose only purpose is to connect sense impressions without logical contradictions. Then there exists, even inside science, a group of men (led by a distinguished astronomer)

who claim to be able to deduce the whole system of natural laws from *a priori* principles without recourse to experience. There are, on the other side, the Marxist philosophers, who consider truth as a deception, and scientific results as a function of social conditions, depending on them and changing with them.

This is, of course, not the place to argue about these different views. I think that modern psychology (doctrine of forms, *Gestalten*), together with the mathematical notion of invariants, is a sufficient basis for dealing with these more or less abstruse ideas, and for developing a realistic and empirical philosophy, the only one which a working scientist can accept.*

The point which matters here is this: If science is only a tool of mankind in producing material progress, there is no need for philosophy. Trained specialists would be all we need. Then man would be just an animal, but less happy because of the complication of his life: working in order to feed, and feeding in order to work. The cow in the meadow simply feeds and is happy. But the cow is in eternal sleep. Man is in a process of awakening, still wrapt in dreams, but here and there catching a glimpse of the real world surrounding him. These clear moments distinguish him from the brute, and they give him his dignity. Humanity is a collective enterprise, and the individual shares in the common stock according to his powers. Scientists are custodians of collective property, just as artists, and learned men of all descriptions, are in their domain. But there is nobody who has not some part in it.

Science itself is the condensation of some of the facts, snatched by the awakening sleeper, facts bewildering by their strangeness, but exciting by their greatness. There is no doubt that the problem of the stellar universe has fascinated the human mind from the earliest times, and that the stages of its development—Babylon, Egypt, Ptolemy, Copernicus, Galileo, Newton, Einstein—are marks in the history of man, coinciding with high tides of culture in other domains. The characteristic feature of our modern times is the analogous development of the Microcosmos, the world of atoms and sub-atoms; though it does not appeal to the average man's imagination in the same way as the stellar universe, it has in common with it the power of creating new conceptions.

This is a most important point in the bearing of science on general education. Between the vagueness of every-day notions (about objects presented by life) and the super-clearness of pure mathematics (about objects invented by the mind) there is the system of theoretical science, i.e. of conceptions formed by interpreting systematic experiments on natural objects. The formation of the Newtonian concept of mass, or of the Daltonian concept of atom, or of the Faraday-Maxwell idea of field would be, I consider, most instructive for everybody. Even an analysis of the meaning of " length " or " distance " has a high pedagogic value; for it reveals, by a simple example, the continuous change

* See my Edinburgh Inaugural Lecture, " Some philosophical aspects of modern physics " (*Proc. Roy. Soc. Edin.* **57**, 1 (1937).

and adaptation of a notion to the empirical situation. Comparison of length involves the idea of transportable rigid bodies. It can be performed practically only in a very limited number of cases, and has to be replaced by indirect methods in the domain of very small and very large bodies: in the micro-domain we use wave-lengths of light, of x rays, of electronic waves, or even less direct evidence such as scattering of electrons and α-particles, and so on; in the macro-domain we use trigonometric triangulation, the velocity and other properties of light, empirical laws about the correlation of luminosity and spectra of stars, the periods of Cepheid variables and other phenomena. This is sufficient material for explaining the involved nature of even the simplest and apparently evident notions, when considered from the standpoint of empirical science (the "operational" method of definition, according to Bridgeman).

By following this method it should be possible to convey to the student at least some idea of modern generalizations: statistical mechanics, relativity and quantum theory, which touch the old and famous philosophical problems concerning space and time, matter, cause, chance, and probability. At present these are left to the philosophers proper; but I doubt whether students, lacking any knowledge of the actual procedure of science, can grasp from them more than empty words. The fight against word-fetishism should be one of the foremost aims of teaching science.

The practical question of how such a programme can be fitted into the frame of our university organization is difficult to answer, particularly for me, as I myself have not had a British education and have known the system only for a relatively short period. But I have had long experience in Germany, and some in the United States.

In Göttingen there was a regulation according to which every professor had to deliver a *publicum*, a public (gratis) lecture (one hour weekly), for students of all faculties at least every sixth semester (third year). Though the rule was generally neglected, we mathematicians and physicists took it seriously, and these lectures were given every Saturday morning, to crowded auditories, on all kinds of fundamental questions of mathematics, physics, and the borderland between other sciences and philosophy. An institution like that is difficult to introduce here, where all courses have a fixed syllabus and time-table, and I doubt whether a Saturday morning lecture, even if gratis, could compete with sport and other week-end activities. But there are other days and hours in the later afternoon or evening.

This, however, would be hardly sufficient to raise the level of science in general education. I feel that there should be a regular course in Natural Philosophy for the ordinary degree, parallel to that in Moral Philosophy and in Metaphysics, or that part of the time fixed for these subjects should be devoted to an introduction to Natural Philosophy. The aim of the course should not be the acquisition of scientific technique of any kind, and the working out of

examples should be restricted to the very simplest ones which are essential for testing the student's understanding. The stress should be laid on the experimental methods and the theoretical interpretation, on the forming of exact concepts about observations, and on the fundamental results, as I have tried to outline above.

Proceeding now to the second group of students, the " professionals ", the only ones who, under the present system, are taught some theoretical physics, I can, from my Edinburgh experience, suggest some possible reforms, although the practical realization of them is at least as difficult as in the former case.

An introduction to theoretical physics should lead the student to a level where he has a glimpse of the front line of to-day's research, and which should enable him to study modern branches of his subject without difficulty. Such an introduction must, therefore, comprise the following courses (apart from the necessary preparatory subjects, i.e. experimental physics and some parts of mathematics):

1. Statics and dynamics of systems of mass points and rigid bodies;
2. Elasticity, hydrodynamics and aerodynamics;
3. Classical thermodynamics, including the elements of the theory of chemical equilibrium;
4. Electricity and magnetism, with an outline of relativity;
5. Classical theory of the electron, electromagnetic waves and optics;
6. Statistical mechanics, based on the kinetic theory of gases; elements of quantum statistics, with application to solids and radiation;
7. Elementary quantum theory and structure of atoms and molecules;
8. Introduction to wave mechanics.

The time available for this programme is not sufficient. Here in Edinburgh we have students working partly for the (honours) M.A., partly for the B.Sc., and the total time available is, for most of them, four years, and is, in exceptional cases, reduced to three. (The same holds, *mutatis mutandis*, in most universities —Cambridge has to be considered separately.) The first year is occupied by experimental physics and mathematics (apart from a short course in very elementary mechanics, necessitated by the unequal standard of the schools). The second year is devoted to statics and elementary dynamics (i.e. mainly such problems as can be reduced to two spatial dimensions), the third to higher dynamics, electricity and magnetism, and a course on elastic vibrations and hydrodynamics. There are no special lectures for the fourth-year students, who repeat the third-year course and do some reading. The reason for this practice is, partly, the occurrence of exceptionally gifted students who are given the opportunity of going through the whole set of lectures in three years and, partly, the smallness of the staff.

This programme covers only items 1, 4 and a part of item 2 of my list. It can hardly be called a training in theoretical physics. As a matter of fact, it

was not so called, but was designated an honours course in "Applied Mathematics". I have changed this title, at least for the B.Sc. degree, to "Theoretical Physics", but without being able, so far, to reform the syllabus appreciably. I shall not, however, discuss the practical means of reform, but simply state what, in my opinion, should be its aims.

In Göttingen there was no fixed syllabus and there were no prescribed courses. Lectures were simply offered, and the student could attend which he pleased. The average duration of study was about four years, or eight semesters. Subtracting one year for mathematical preparation, there remained six semesters for accommodating the eight subjects of my list; this was impossible. Two of them had, therefore, to be eliminated or merged into other lectures. This was done with item 2, elasticity and hydrodynamics, and with item 8, wave mechanics. The latter is of less importance; for, owing to the length of the semester compared with our term (14 weeks against 10), wave mechanics could be dealt with in the general lectures on quantum theory and atoms (item 7). On the other hand, the handling of mechanics of continuous media raises a question of principle, connected with an essential point of nomenclature. This subject has, under modern conditions, acquired a tremendous practical importance. It is the foundation of all engineering; aerodynamics in particular is the science of flying, and, as such, part of the country's defence. No wonder that these two branches have grown out of all proportion to what they were when the statutes of our faculties were drawn up; and the part which can be taught in a one-term course of university lectures is ridiculously small. In view of this fact there was in Göttingen a special department, called "Applied Mathematics and Mechanics", in which the real thing was taught: not the theory of non-existent ideal solids and liquids, but the modern theories of elastic and plastic solids, viscous fluids, vorticity and surface layers, and applications to practical conditions. Only the fundamental concepts were taught in the general course of theoretical physics, and these could be easily fitted into analytical mechanics (item 1) and thermodynamics (item 3).

I think such a separation is unavoidable. This statement is not in contradiction with my regret that the unity of science is endangered by its subdivision into separate branches. For here we are not concerned with general education, for which this unity is essential, but with the efficiency of teaching special subjects. The theory (and practice) of elasticity and hydrodynamics is great and important enough for a department of its own, forming a very desirable link between pure science and engineering. It is not concerned with the discovery of new laws of nature, but with the application of given laws. This is not a minor job, but a different one, needing a man's full strength. There is to-day hardly anybody who could be called an expert in both fields.

Whether the name of this new department should be "Applied Mathematics" or "Applied Mechanics" or (as in Göttingen) a combination of them, may be left to those responsible. This question of naming is not unimportant,

and I confess that I should prefer another title, say " Mechanics of Engineering ". For there are very many sciences applying mathematics to their special problems; besides theoretical physics there is astronomy, theoretical chemistry, crystallography, statistics (including such divergent subjects as insurance systems and the theoretical aspects of genetics), meteorology and so forth. Each of them can claim an independent existence and rightly refuse to be considered as an appendix to mathematics. The custom of calling mechanics and theoretical physics by a title suggesting that the centre of interest lies in the application of mathematical skill ought to be abandoned. Mathematics is a tool for the theoretical physicist, as glass blowing and vacuum techniques are for the experimentalist. But you would not call the Cavendish "Applied Glass Blowing ".

Returning now to the teaching of theoretical physics, let us suppose that the suggested segregation of engineering mechanics has been made, as is, in fact, the case in Central Europe and many places in America. Let us further assume that the staff is large enough to provide a separate course of general education, as described above. Then for the " professionals " there is available one preparatory year for experimental physics and mathematics, and three years (nine terms) for theoretical physics, supplemented by mathematical and other lectures.

We theoretical physicists should also have a say in the framing of the first-year syllabus of the experimental and mathematical lectures. We need a lecture course on physics with demonstrations to cover the whole field. Practical courses are very important, for one has to fight the inclination of many students to consider the theory as a pleasant mathematical game or entertainment without relation to reality. An ideal syllabus should therefore include experimental courses during the whole period of study, even if it were only for a few hours a week. In Göttingen I had started an arrangement which seemed to prove very satisfactory. One of the lecturers of the experimental physics department (Dr. Kuhn, now at Oxford) gave, as a supplement to my lecture on theoretical optics, a demonstration lecture, of one hour a week, in which he showed the fundamental experiments (interference, diffraction, dispersion, absorption etc.) on which the theory is built. The continuation of this attempt was prohibited by political events. I think that such supplementary experimental lectures could be given for all subjects. Thermodynamics, for instance, would be a most attractive subject on which to base such a course. I have heard that similar arrangements have been made in American universities.

Proceeding now to mathematics, it is obvious that theoretical physics cannot be started without some knowledge of analytical plane geometry and some calculus. These must therefore be given in the first year. Provided this condition is fulfilled, there is some freedom in the order of the mathematical subjects needed in the following years; the best sequence would be this: More advanced calculus, in particular three-dimensional integrals; analytical and, a little, differential geometry in space, connected with an introduction to linear

algebra and vector analysis; combinatorial analysis; elementary differential equations, for one and more independent variables. Very desirable, too, is an introduction to eigenvalue problems and to group theory.

A special word must be said about vector analysis. It would be a great help if it could be given as early as possible by the mathematicians; and it should be systematically used in theoretical physics from the very beginning. The notation is today sufficiently settled; the only disputable point is the vector and the scalar product, but I think that the symbols $\boldsymbol{a} \,.\, \boldsymbol{b}$ and $\boldsymbol{a} \wedge \boldsymbol{b}$ (recommended by Chapman, Hartree, Milne and others) are now generally accepted. In the third year, students should learn the tensor calculus; as there is no general agreement about tensor symbols I prefer the suffix notation with Einstein's convention (omitting summation signs). We should be grateful if the mathematicians would provide for all these tools of our business.

The second year should be entirely devoted to mechanics: the first term to statics, the two following terms to dynamics, including the fundamental concepts of the treatment of continuous media.

The following two years, six terms, are just sufficient for the six remaining subjects of my list: thermodynamics, electro-magnetism, electronics and optics in the third year; statistical mechanics, quantum theory and wave mechanics in the fourth. The order can, of course, be changed to a certain extent, according to circumstances, although the general line is given by the matter itself.

The shortening of the time allotted to mechanics (as compared with the present usage in Edinburgh, as described above) can be justified by the fact that dynamical questions crop up in many of the other subjects, in thermodynamics, electro-magnetics, optics (dispersion) and atomic structure. I am strongly in favour of a reduction of the course in statics which, in its present form (as represented by many textbooks), is a preparation more for engineering or architecture than for physics. I cannot discover much educational value in the solution of tricky questions concerning amazing combinations of pulleys, spheres, frictionless joints and weightless bars. Their only purpose is to supply nice problems for tutorials and examinations. Certainly the collections of such problems in the textbooks are admirable in their own way. They are well suited for exercising a certain kind of intelligence. My objection is, that we are not concerned with intelligence tests, but with teaching a living and growing science which should be represented, from the very beginning, not as a dull practice but as an intellectual adventure.

This question of problems for tutorials and examinations appears also in the higher branches, particularly in electricity, with its old Maxwellian tradition. I have come across a great many excellent examples, instructive and amusing at the same time. Many of them are, however, too remote from reality. To give one example—well, perhaps not! I think it is clear which kind of

problems I mean: they are just "applied mathematics", using very special theorems and tricks, and leading nowhere, apart from the examination marks.

The reverse side of this criticism is that I myself am completely unable to suggest a collection of questions which I should prefer.* The making and solving of such questions is a technique to be properly learned in youth, and I never had the opportunity. For in Germany this side of the training was neglected—badly neglected. I quite appreciate its value; each technique acquired in youth is an advantage in later life. It might have saved me much trouble, time and thought—the last, however, I do not regret.

This brings me to another point concerning the methods of teaching. They seem to me to lay too much stress on knowledge and technique, and to neglect the play of imagination. Science is not formal logic—it needs the free play of the mind in as great a degree as any other creative art. It is true that this is a gift which can hardly be taught. But its growth can be encouraged in those who already possess it.

This point brings me to the third group of students, the small minority of those who are above the average and wish to devote their lives to our science. These are individuals who conform to no rule. Heisenberg almost failed in the oral examination for his doctorate, and that on a question concerning the resolving power of a grating—precisely the problem which led him, a few years later, to his celebrated relations of indeterminacy. Dirac was an engineer; he changed over to theoretical physics and created immediately a new kind of mathematics for his form of quantum mechanics. Genius cannot be guided. But opportunity can be offered for shortening and easing its way.

There are two new foundations dedicated to this purpose, each designated "Institute for Advanced Studies", one in Princeton, U.S.A., the other in Dublin, Eire. Both have mathematical physics as a main subject (in Dublin it is at present the only one), an acknowledgment of its fundamental position between the abstract constructions of the mind (mathematics) and the empirical sciences. These institutions provide the most efficient teaching for a small selection of gifted students.

* I am reluctant to recommend textbooks, as there are too many which I do not know. There is one book in English which contains the whole matter in one volume, well written, cleverly selected, and with many problems: J. C. Slater and N. H. Frank, *Introduction to Theoretical Physics* (McGraw-Hill, 1933). There are two similar foreign books, also in one volume, both very good. The first one, by G. Joos (Göttingen), was originally written in German: *Lehrbuch der theoretischen Physik* (Akademische Verlagsgesellschaft), 3rd ed. 1939, and has been translated into English (by Freeman): *Theoretical Physics* (Blackie and Son, 1934); it contains a good collection of examples. The second one, by R. Fürth (Prague, at present in Edinburgh), is written in German: *Einführung in die theoretische Physik* (Springer, Wien, 1936). Then there are Max Planck's lectures: *Einführung in die theoretische Physik* (Hirzel, 1931), translated into English (by Brose): *Introduction to Theoretical Physics* (Macmillan, 1932, 33), in five small volumes which, although not quite up to date, are a good example for the arrangement of the whole subject in separate courses. There are plenty of textbooks on the different branches, but I am unable to recommend a selection of these, apart, of course, from my own. In fact, my little book *Atomic Physics* (Blackie and Son, 1937) covers the subject of the last-year courses and may be useful.

To some extent Oxford and Cambridge play the same part in Britain, although their main task is the general education of those who are expected to become the leaders of the nation. I shall not touch the question whether special research institutes ought to exist besides these schools of general learning. But I have to mention one point connected with Cambridge.

Universities like my own are restricted with respect to advanced teaching by the number of the members of the staff. It is desirable that each teacher, professor or lecturer, should spend at least one hour weekly on an advanced lecture or seminar beyond the syllabus. But it is often difficult to spare even this little time. Hence a student who wishes to do research has not much opportunity to extend his knowledge at his own university, and he prefers to go to Cambridge, using the scholarships which are provided for the top men of the classes. Even if a local research school exists (such as I have tried to create here, mainly with students from foreign countries), it is known that a Cambridge degree is highly valued and opens up many more opportunities than any other degree. This predominance of the oldest universities is natural and sound, but in order to be effective as a means of selecting the best it must be permanently challenged. Each other university should have the ambition to be just as good, at least in some subjects. Departments with flourishing research schools ought to be encouraged and provided with means. Theoretical physics is a very " cheap " subject : paper, ink, pen, a few books and a calculating machine are all we need ! A sufficient staff and a number of research scholarships attached to the department would enable it to keep good students from going elsewhere.

THE THEORY OF THE SPECIFIC HEAT OF SOLIDS

By M. BLACKMAN,
Imperial College, London

§ 1. INTRODUCTION

THE specific heat of a substance under given external conditions (denoted by x) is defined by $c_x=(dQ/dT)_x$, i.e. the ratio of the heat, added to a gram of the substance, to the temperature rise. In most cases the specific heat is measured at constant pressure (c_p) and at constant volume (c_v). When experiments yield c_p, as is the case in solids, this can be converted to c_v by means of the general thermodynamical relation

$$c_p-c_v=9\alpha^2VT/\kappa,$$

where α is the coefficient of linear expansion, V the specific volume, κ the compressibility, and T the absolute temperature.

This conversion is necessary because the theory gives specific heats, or rather heat capacities, per gram molecule at constant volume. It is usual, in theoretical considerations, to ignore the expansion of the solid and the consequent changes in elastic properties; this neglect involves only small errors (probably of a few per cent (Kellermann, 1941)), and can be remedied if the accuracy of the theory should warrant it.

Historically, the theory of heat capacities falls into two distinct parts: (*a*) classical theory, which yielded a constant heat capacity per gram molecule of $3\boldsymbol{R}$(where $\boldsymbol{R}$ is the gas constant per gram molecule), and (*b*) quantum theory, due in the first instance to Einstein (1907), and given a more complete form by Debye (1912), Born and v. Kármán (1912), and Born (1923). The full implications of the crystal theory of Born have, however, been worked out only recently (Blackman, 1935-7; Kellermann, 1941), and this has thrown some light on the more detailed properties of the heat capacity shown by experiments at low temperatures.

From the point of view of classical theory the rather complex problem of dealing with a large number of particles bound together to form a solid presented little difficulty. Since the solid is stable, the heat energy is simply the energy of a set of vibrating particles, the average value being $3N\boldsymbol{k}T$, (N is number of particles, $\boldsymbol{k}$ Boltzmann's constant and T the absolute temperature). Hence $C_v=3N\boldsymbol{k}$, which is independent of temperature. This sufficed to explain the

well-known law of Dulong and Petit—the atomic heat capacity at constant pressure is equal to 6·2 for a large number of substances. It did not explain the exceptions to the Dulong and Petit law, the low values for some substances, e.g. diamond at normal temperatures, or the low values obtained at liquid-air temperatures for substances obeying the Dulong and Petit law at high temperatures (Behn, 1898).

The solution to this difficulty was given in a fundamental paper by Einstein (1907), who applied the (then) new quantum theory of Planck to the motion of the particles of a solid, i.e. he replaced the classical mean energy for a linear oscillator $\bar{\epsilon} = \boldsymbol{k}$ by the expression $\bar{\epsilon} = \boldsymbol{h}\nu/\{\exp.(\boldsymbol{h}\nu/\boldsymbol{k}T) - 1\}$, where ν is the frequency of oscillation and $\boldsymbol{h}$ is Planck's constant. The general problem was idealized by assuming every particle in the solid to vibrate with a single frequency, and the heat capacity for N particles then assumes the form

$$C_v = 3N\frac{d}{dT}\left(\frac{\boldsymbol{h}\nu}{e^{\frac{\boldsymbol{h}\nu}{\boldsymbol{k}T}}-1}\right) = 3N\boldsymbol{k}\left(\frac{\boldsymbol{h}\nu}{\boldsymbol{k}T}\right)^2\frac{e^{\frac{\boldsymbol{h}\nu}{\boldsymbol{k}T}}}{\left(e^{\frac{\boldsymbol{h}\nu}{\boldsymbol{k}T}}-1\right)^2} = 3NE\left(\frac{\boldsymbol{h}\nu}{\boldsymbol{k}T}\right).$$

This formula fitted the experimental data for diamond very well (as far as these were known, i.e. down to liquid-air temperatures). Though the fundamental difficulties were now solved, the details of the motion of the particles were not solved, and no very accurate agreement with experiment could be expected.

As the Einstein theory indicates, it is essential to know more about the frequencies of vibrations of the solid. The general theory of the motion of a dynamically stable system of coupled particles shows that one can resolve the motion into a superposition of number of independent oscillations of the system (" normal vibrations "), there being $3N$ of these (more strictly, $3N-6$), if N is the number of particles in the system. Each of these normal vibrations will have a definite frequency. If these frequencies are known, the total heat capacity can immediately be determined, since it is simply the sum of the contributions due to the individual vibrations,

$$C_v = \sum_{\nu} E(\boldsymbol{h}\nu/\boldsymbol{k}T).$$

A strict proof of this relation can be given, using quantum statistics (*vide*, for instance, R. H. Fowler, ***Statistical Mechanics***). Quantum theory adds the zero-point energy, $\frac{1}{2}\boldsymbol{h}\nu$, to the expression used by Einstein for the mean energy of a linear oscillator, but the expression for the heat capacity remains unaltered.

The above summation can be turned into an integral if we introduce a function $\rho(\nu)\Delta\nu$ giving the number of frequencies between ν and $\nu+\Delta\nu$; $\rho(\nu)$ represents the vibrational spectrum. In terms of $\rho(\nu)$, the heat capacity becomes

$$C_v = \int_0^{\nu_m} \rho(\nu)E(\boldsymbol{h}\nu/\boldsymbol{k}T)d\nu.$$

As regards the general features of the spectrum, one knows, firstly, that there is a lower limit which depends on the properties and dimensions of the crystal,

but which is sufficiently low to be put equal to zero for the purpose of calculating the specific heat; secondly, there is a definite upper limit (ν_m)—the highest frequency with which the solid can vibrate—which depends on the binding forces in the solid and on the masses of the particles taking part in the vibration.

It follows that the heat capacity must be a continuously increasing function of the temperature, since the function E has this property; also, since $E(h\nu/kT) = k$ when $h\nu \ll kT$, it follows that we obtain the classical form for C_v when $h\nu \ll kT$. The accurate description of specific-heat phenomena depends on a knowledge of the spectrum. It is, for this reason, instructive to consider the various theoretical approximations to the spectrum implicit in the different theories of specific heat.

Einstein's theory replaces the whole of the spectrum by a single line at a definite frequency (chosen to fit the experimental data), which will usually be in the neighbourhood of the maximum frequency.

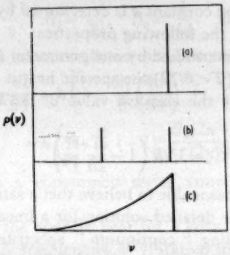


Figure 1. Vibrational spectra: (*a*) Einstein; (*b*) Nernst-Lindemann; (*c*) Debye. The same value has arbitrarily been chosen for the characteristic frequency in each of the cases (*a*), (*b*) and (*c*).

Einstein's specific-heat function proved rather unsatisfactory at low temperatures, and an empirical formula was put forward by Nernst and Lindemann (1911) in order to fit the results. This consisted in replacing the spectrum by two "lines" of equal weight, one at a frequency ν and one at $\nu/2$. This specific-heat function was very successful over a wide region of temperatures, though it failed at the lowest temperatures. It is rather a curious point that a theoretical justification for the Nernst-Lindemann spectrum is provided by recent work on the spectrum of cubic crystals (see figures 3, 4, 5).

The next step in the development of the theory was made almost at the same time by Debye (1912) and by Born and v. Kármán (1912). Debye showed that the function $\rho(\nu)$ was proportional to ν^2 in the region of low frequencies, where the theory of the solid could be replaced by that of an isotropic elastic continuum. An approximation to the spectrum was then obtained by continuing the ν^2 function up to a maximum frequency ν_D and cutting off the spectrum sharply at this point

(figure 1). The cut-off ν_D is chosen so as to give the correct total number of vibrations ($3N$), i.e.,

$$\rho(\nu)=a\nu^2, \quad 0<\nu<\nu_D \quad \text{and} \quad \int_0^{\nu_D} a\nu^2 d\nu = 3N,$$

with
$$a=3V\frac{4\pi}{3}\left\{\left(\frac{(1+\sigma)\kappa\rho}{(1-\sigma)}\right)^{3/2}+2\left(\frac{2(1+\sigma)\kappa\rho}{3(1-2\sigma)}\right)^{3/2}\right\},$$

where V = volume of solid, ρ = density, κ = compressibility, σ = Poisson's ratio. From this it follows that

$$\begin{aligned} C_v &= a\int_0^{\nu_D}\nu^2 E(h\nu/kT)d\nu \\ &= 3Nk\left\{12\cdot\frac{T^3}{\theta^3}\int_0^{\theta/T}\frac{\zeta^3}{e^\zeta-1}\,d\zeta-\frac{3\theta/T}{e^{\theta/T}-1}\right\} \\ &= 3ND(\theta/T) \quad \text{with} \quad \theta=h\nu_D/k. \end{aligned}$$

The value of ν_D can be calculated from the elastic constants of the elastic continuum, since the value of the constant a is determined by these elastic constants.

The Debye formula has the following properties:

(*a*) the specific heat is determined by one parameter (θ/T);

(*b*) at low temperatures ($T<\theta/12$) the specific heat is proportional to T^3;

(*c*) at high temperatures the classical value of $3Nk$ is approached in the form

$$C_v=3Nk\left(1-\frac{1}{20}\frac{\theta^2}{T^2}\right).$$

As to the spectrum, it was reasonable to believe that a satisfactory approximation had been obtained, since the detailed solution for a linear chain showed a great similarity to the corresponding "continuum" spectrum—the main difference being the heaping up of vibrations in the neighbourhood of the maximum frequency. It was thought that this effect would be less prominent in the three-dimensional case.

Born and v. Kármán discussed, in their first paper, the calculation of the frequencies of the vibrations of a simple cubic lattice. The detailed spectrum was, however, not worked out at that time and an approximate spectrum was constructed in close analogy with the spectrum of the linear chain. The distribution function $\rho(\nu)$ has the form (ν_0 being the maximum frequency)

$$\rho(\nu)=\frac{72N}{\pi^3\nu_0}\frac{(\sin^{-1}(\nu/\nu_0))^2}{(1-\nu^2/\nu_0^2)^{\frac{1}{2}}},$$

which for small values of ν/ν_0 reduces to the ν^2 form. Both the spectrum and the resultant specific-heat function are rather more complicated than the corresponding Debye functions, without being superior in representing the results of experiments. The Born-v. Kármán function has accordingly receded into the background, but the general theory on which the approximate function was built up has become a part of the lattice theory of vibrations, which forms the basis of much of the recent work on the specific heat of crystals.

§2. EARLY WORK ON THE LATTICE THEORY OF SPECIFIC HEAT

The work of Born and v. Kármán was followed by a very interesting calculation by Thirring (1913, 14) on the specific heat of the cubic lattices dealt with by the former authors. Thirring showed that it was possible to work out the specific heat of these lattices at moderately high temperatures without working out the spectrum explicitly, and without any further approximations. The method is based on the following idea. If one writes the mean energy of a linear oscillator in the usual form,

$$\bar{\epsilon} = \tfrac{1}{2}h\nu + kT\frac{h\nu/kT}{\text{exp.}\,(h\nu/kT)-1},$$

and expands in powers of $h\nu/kT$, one obtains

$$\bar{\epsilon} = kT\left(1 + \sum_{n=1}^{\infty}(-1)^n \frac{B_n}{(2n)!}\left(\frac{h\nu}{kT}\right)^{2n}\right),$$

where B_n are the Bernoulli numbers; the first four have the following values:

$$B_1 = \frac{1}{6}, \quad B_2 = \frac{1}{20}, \quad B_3 = \frac{1}{42}, \quad B_4 = \frac{1}{30}.$$

The heat capacity of the linear oscillator then becomes

$$C_v = \frac{d\bar{\epsilon}}{dT} = k\left(1 - \frac{B_1}{2}\left(\frac{h\nu}{kT}\right)^2 + \ldots\right)$$
$$= k\left(1 + \sum_{n=1}^{\infty}(-1)^n \frac{B_n(2n-1)}{(2n)!}\left(\frac{h\nu}{kT}\right)^{2n}\right).$$

If $h\nu/kT < 2\pi$, this is a convergent series containing terms depending on ν^2, ν^4, ν^6, etc.

In order to obtain the specific heat for a lattice, it is necessary to sum the expression for all the $3N$ frequencies of vibration of the lattice, i.e.,

$$C_v = k\left(3N + \sum_{n=1}^{\infty}(-)^n \frac{B_n(2n-1)}{(2n)!}\left(\frac{h}{kT}\right)^n \sum \nu^{2n}\right).$$

Now the Born-v. Kármán treatment of the vibrations of a crystal lattice gave an expression for the frequencies in the form of a determinant involving ν^2. It is a relatively easy matter to work out $\sum \nu^{2n}$ from this determinant for small values of n, as was shown by Thirring, and so it was possible to work out fully the implications of the theory of the Born-v. Kármán model. Results were obtained for NaCl, KCl, FeS_2 and CaF_2 (Thirring, 1914) as exemplified in table 1.

Table 1

C_p (NaCl)			C_p (KCl)		
$T°$ K.	Observed	Calculated	$T°$ K.	Observed	Calculated
83·4	3·75	3·71	86	4·26	4·29
81·4	3·54	3·67	76·6	4·11	3·95
69·0	3·13	3·09	70·0	3·79	3·66
67·0	3·06	3·02	62·9	3·36	3·30

Thirring's method is completely general and can be applied to any crystal; it has hardly received the prominence it deserves, especially as it was, for about twenty years, the only calculation of specific heats which did not involve the usual approximations which obscure the whole of the effect of a crystal, i.e. a crystal as distinct from a continuum (in the Debye sense).

The development of the theory of specific heats from this point onwards was concerned with the mathematical rather than with the physical side. The general theory of the vibrations of a crystal was worked out by Born (see Born, 1923) following the earlier work of Born and v. Kármán (1913). It was shown that the vibrations of a crystal could be divided into $3s$ groups, where s is the number of particles in the unit cell of the crystal. Of these, three represent low-frequency (long wave-length) vibrations which in the limit become identical with the vibrations of an elastic continuum. These are termed *acoustical* vibrations and the remaining $3s-3$ *optical* vibrations, the frequencies of these latter vibrations being, as a rule, of the order of optical frequencies ($\sim 10^{13}$).

The general equations determining these frequencies were worked out for the case of central forces, but the determination of the spectrum was (and in most cases still remains) far too complex and laborious to be attempted; the specific heat was therefore given only in an approximate form. In the simplest formulation, each of the three acoustical groups of vibrations was represented by a Debye spectrum and the optical vibrations by single-line (Einstein) spectra; the specific heat then becomes

$$C_v = \sum_{j=1,2,3} D(\theta_j/T) + \sum_{j=4}^{3s} E(\theta_j/T).$$

The functions D and E are suitably normalized Debye and Einstein functions. In the case of monatomic crystals, in which the unit cell contains one particle, the above expression reduces to the terms with D only.

The values of θ_1, θ_2, θ_3 can be calculated from the elastic constants; in some cases the values of the optical frequencies can be obtained from absorption bands in the solid, though these are not necessarily accurate representations of the positions of the maxima in the spectra. The calculations can therefore be carried out in only a few cases, mainly for crystals of the alkali-halide type. These were made by Försterling (1920 b), who found good agreement with experiment in the cases of KCl, NaCl and CaF_2.

It should, however, be noted that in none of these is the agreement with experiment very much better than with a Debye formula, and that the initial approximations made to the spectrum are probably no closer in these cases than in the assumption of a Debye spectrum. The approximations involved can be seen in figure 2, in which the spectrum of a simple cubic lattice (Blackman, 1937) is given together with the representation as the sum of the three Debye spectra.

The main use of the general formula of Born lies in its indication of the way in which an approximate formula can be built up in the case of complicated crystals; this serves a very useful purpose in calculating the entropy of such

crystals. An outstanding example of a representation on these lines is the specific-heat curve of benzene (Andrews, D. H.; *cf.* Eucken, 1929).

Some calculations on simple crystals at high temperatures were also made by Försterling (1920 a, c) using the method of Thirring. The equations for the frequency of the vibrations were, however, those of the continuum and not those of the crystal lattice, so that there is no real advantage as compared with the Debye theory.

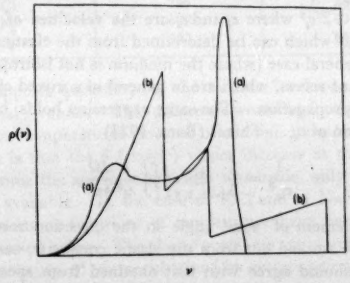


Figure 2. Vibrational spectrum of a cubic lattice, curve (*a*), together with the representation of the spectrum as a sum of three Debye spectra, curve (*b*), obtained from the low-frequency end of the spectrum.

A rather different approach to the representation of the specific heat of complicated crystals (which show a deviation from Debye theory) was made by Lewis and Gibson (1917). The specific-heat curve was fitted with a function $D(\theta^n/T^n)$, where n is an empirical constant. Some success was obtained, as is to be expected with two parameters instead of one, but there does not seem to be any theoretical justification for the form of the function.

§ 3. LATTICE THEORY AND SPECIFIC HEATS AT LOW TEMPERATURES

Although the general theory of crystals did not at first provide a detailed description of the vibration spectrum, Born (1923) showed that when the wavelength of the vibration was long (compared with the crystal spacing) the crystal behaved like a continuum, and further (from the general theory of normal vibrations), that the spectrum followed the ν^2 law. On this point the theory agrees with and extends the Debye theory, and in particular it follows that the specific heat must follow a T^3 law at " sufficiently " low temperatures. One of the points left open in the general theory is a mathematical description of the important adverb " sufficiently ". If the equations for the frequencies are known, the temperature below which the T^3 law holds can be worked out from case to case, but it will depend very much on the forces between the particles in the crystal.

Debye (1912) showed that, in the case of an elastically isotropic medium, the heat capacity had the form (if $T<\theta/12$),

$$C_v = \frac{12\pi^4}{5} Nk \frac{T^3}{\theta^3},$$

where $\theta = \frac{h}{k} c_m \left(\frac{3}{4\pi} \cdot \frac{N}{V}\right)^{\frac{1}{3}}$; c_m is a mean velocity of elastic waves given by the relation $3/c_m^3 = 1/c_l^3 + 2/c_t^3$ where c_l and c_t are the velocities of the longitudinal and transverse waves which can be determined from the elastic constants of the medium. In the general case (where the medium is not isotropic) the velocities of the three groups of waves, which are in general of a mixed character, depend on the direction of propagation. The same expression holds, however, with an appropriate definition of c_m. This is (Born, 1923)

$$\frac{1}{c_m^3} = \frac{1}{3} \sum_{j=1,2,3} \iiint \frac{1}{c_j^3} \frac{d\Omega}{4\pi},$$

where $d\Omega$ is the element of solid angle in the direction associated with c_j. This quantity can be worked out from the elastic constants, and the value of θ found in this way should agree with that obtained from specific-heat data at low temperatures. It is of course necessary that the elastic data used should also be low-temperature values, but in general the extrapolation to low temperatures can be made without serious error.

A number of investigations have been concerned with the calculation of the value of c_m, as the direct evaluation is rather laborious. In the case of cubic crystals which are nearly isotropic, a simple formula has been given by Born and v. Kármán (1914), namely,

$$\frac{3}{c_m^3} = \rho^{3/2} \left\{ \frac{2}{c_{44}^{3/2}} + \frac{1}{c_{11}^{3/2}} + \frac{3}{5} \left(c_{12} - c_{11} + 2c_{44}\right) \left(\frac{1}{c_{44}^{5/2}} - \frac{1}{c_{11}^{5/2}} \right) \right\},$$

where c_{11}, c_{44}, c_{12} are the three elastic constants of the cubic crystal (in Voigt's notation), ρ the density, and it is assumed that $(c_{12}-c_{11}+2c_{44})/(c_{11}-c_{44})$ is small compared with unity. (The original formula of Born and v. Kármán includes a factor of 4π, which seems to be an oversight.) A very useful numerical method is due to Hopf and Lechner (1914); the method is applicable, in principle, to all cubic crystals, and can be extended to other types of crystals. The general applicability of the numerical method has, however, tended to obscure the usefulness of the Born-v. Kármán formula. For example, the case of tungsten has been treated (Honnefelder, 1923) by the, by no means quick, numerical method, whereas the value of c_m could have been worked out immediately as the elastic constants for tungsten satisfy the condition for isotropy, i.e. $c_{12}-c_{11}+2c_{44}=0$. In the case where $(c_{12}+c_{44}) \ll c_{11}$ an explicit formula can be given for the θ value (Blackman, 1935 d).

The first comparison of the theoretical values of c_m (or θ) with those deduced

from experiments at low temperatures provided quite interesting results (Hopf and Lechner, 1914; Eucken, 1929) which are shown in table 2.

Table 2

Crystal	θ (elastic data)	θ (thermal data)
KCl	230	224
NaCl	296	286
CaF_2	499	470
FeS_2	682	645

The agreement appears quite fair, but the " elastic " values are all slightly higher than the " thermal " values. This difference would be further accentuated if the elastic data at low temperatures (instead of at room temperature) were used, as the general rule is that the θ (elastic) values increase as the temperature goes down. At the time the above comparison was made, only a rough estimate of the change was available. In the case of KCl and of NaCl, measurements of the elastic constants as functions of temperature have been made by Durand (1936) and are quoted below. The elastic constants are in units of 10^{11} dyne/cm^2; the values at 0° K. being, of course, obtained by extrapolation.

Table 3

KCl				NaCl			
T° K.	c_{11}	c_{44}	c_{12}	T° K.	c_{11}	c_{44}	c_{12}
280	4·09	0·634	0·60	400	4·40	1·234	1·17
180	4·45	0·655	0·60	300	4·82	1·264	1·17
0	4·91	0·669	0·60	0	5·85	1·339	1·17

The increase in the value of c_m can be obtained fairly accurately from the above elastic constants, and is ~5 % for KCl and ~6 % for NaCl. One would expect that the θ values for CaF_2 and FeS_2 would have to be increased by similar amounts. If this is the case, the discrepancy for CaF_2 becomes 11 %; these two values should, according to crystal theory, be identical. The case of CaF_2 is of particular interest as the specific heat at low temperatures follows the T^3 law so well as to make it one of the standard examples of the validity of this law.

Discrepancies of about 10% fade, however, into the background when the comparison is made for some other single crystals (table 4: cf. Eucken, 1929).

Table 4

Substance	θ (elastic data)	θ (thermal data)
Zinc	309	205
Cadmium	180	128
Zincblende	350	270

Here again the elastic data are for room temperature, the thermal data for $\sim 20°$ K. These discrepancies were particularly puzzling, because in the case of zinc the specific heat at low temperatures shows very good agreement with the T^3 law (Clusius and Harteck, 1929).

Two suggestions were made at the time to account for the difference. Eucken (1929) suggested the possibility of large systematic errors in the experimental measurement of the elastic constants. Grüneisen and Goens (1926), who carried out the experiments, derived a "theoretical" correction to the θ (elastic) value of the right order of magnitude. Unfortunately, this correction does not appear to be sound (Eucken, 1929; Blackman, 1935 a), as the two values must agree at sufficiently low temperatures. A certain amount of light was thrown on this point by investigations into the specific heats of lattices, which showed that a T^3 region at low temperatures is no proof that the "continuum" region has been reached, i.e. it is possible for spurious T^3 regions to be found (Blackman, 1935 b). Furthermore, it was found that the criterion of Debye, $T<\theta/12$, was by no means a good criterion of "sufficiently low temperatures". The conclusion is, therefore, that measurements of the specific heat at still lower temperature in the above cases would show a decided change, and that eventually the true T^3 region would be reached; for this region the θ value should fit with that determined from elastic data at low temperatures.

There are measurements of a few substances at helium temperatures, which allow the above conclusions to be tested (table 5). Zinc has been investigated by Keesom and van den Ende (1932), silver by Keesom and Kok (1932) (table 8). In both these cases there is a complication in that the specific heat due to the free electrons becomes important. This is a term proportional to T (cf. Sommerfeld and Bethe, 1933), and the θ value at 4° K. was found by fitting a curve $C_v = aT + bT^3$, to the measurements, it being assumed that b was constant below 4° K. In the case of rocksalt and sylvine there are direct measurements by Clusius and Perlick (*vide* Keesom, 1934) and by Keesom and Clark (1935). The θ (elastic) values for the two last-named substances are due to Durand (1936).

Table 5

Substance	θ (thermal)	$T°$ K.	θ (elastic)	$T°$ K.
Silver	237	4	216	~ 290
Zinc	308	4	305	~ 290
Rocksalt	308	10	320	0
Sylvine	230	3	246	0

The most striking fact is that the very puzzling discrepancy in the case of zinc has practically disappeared. It should also be noted that there is good reason for believing (§ 4) that for KCl and NaCl the θ (thermal) values at the absolute zero are rather higher. The value for silver does, however, seem a little high if the correction to the elastic value, to allow for the difference in

temperature, is not more than 5 %. The subtraction of the electronic term does, however, make the value somewhat uncertain, and it may be as low as 220 at 2° K.

What does seem evident from the above is, that the specific-heat curve should be followed to much lower temperatures than one would have thought likely on the Debye theory. A critical review of the Debye theory does, therefore, appear necessary.

§ 4. COMPARISON OF DEBYE THEORY WITH EXPERIMENT

The Debye theory had remarkable success in representing the variation with temperature of the specific heat of a large and varied assortment of substances.

Table 6

Substance	θ			Reference
	$T\sim\theta$	$T\sim\theta/6$	$T\sim\theta/12$	
Gold	180	172	162	Clusius and Harteck (1928)
Silver	220	210	209	Eucken, Clusius, and Woiteneck (1931)
Copper	315	317	319	Dockerty (1937)
Platinum	230	224	220	Simon and Zeidler (1926)
Lead	88	87	85	Keesom and v. d. Ende (1931); Eucken and Schwers (1923)
Lithium	430	379	356	Simon and Swain (1935)
Sodium	150	159	—	Simon and Zeidler (1926)
Potassium	99	98	—	Simon and Zeidler (1926)
Tungsten	310	305	337	Lange (1924) ; Zwikker (1928)
Molybdenum	378	378	378	Simon and Zeidler (1926)
Magnesium	—	318	329	Clusius and Vaughan (1930)
Cadmium	170	147	129	Lange and Simon (1928)
Zinc	240	211	205	Clusius and Harteck (1928)
Sylvine	230	225	218	Southard and Nelson (1933)
Silver bromide	156	137	118	Eucken, Clusius, and Woiteneck (1931)
Silver chloride	211	170	133	Clusius and Harteck (1928)

It is hardly surprising that it was taken for granted that the underlying assumption (i.e. the general form of the vibrational spectrum) was sound, so long as the agreement was good. This tendency took a rather exaggerated form when it was thought necessary to assume "extra" effects whenever deviations from the Debye theory were observed—an interesting example of what Spengler terms the "canonization" of a physical theory.

The Debye theory was originally derived for monatomic and (elastically) isotropic substances. It has been applied unhesitatingly to all types of crystals. Seeing that the ν^2 law for the spectrum is perfectly general, there does indeed seem no valid reason for discrimination. It should be noted that the lower end of the spectrum—in the isotropic case—is made up of separate contributions

from (*a*) longitudinal, (*b*) transverse waves, and that the relative weights of these can be varied over a considerable range if the two elastic constants of an isotropic medium are allowed to vary. There is, indeed, only one cubic crystal for which the condition for isotropy is perfectly fulfilled—tungsten—and this is one of the cases where an appreciable deviation from the Debye curve is found.

In table 6 the experimental results are shown for a variety of substances. The Debye function gives a unique correlation between the heat capacity per gram atom and the value of θ/T. Hence any specific-heat curve can be represented (as a function of T) by a set of values of θ as a function of T. A perfect fit with the Debye curve would mean a constant value for θ. In the table the values of θ are given in three temperature regions—at relatively high temperatures, at relatively low temperatures (where the T^3 region should start on Debye theory) and at an intermediate temperature.

The conclusion from the table would certainly be that in general the agreement is good, particularly for face-centred cubic crystals. There are deviations for gold and for tungsten of about 10 % in the θ value in the extreme cases, and rather larger deviations for lithium, zinc and cadmium (30 % to 50 %). These three substances are strongly anisotropic elastically, and it might seem that there is a correlation between this " anisotropy " and the deviation from the Debye theory (Fuchs, 1939). Against this may be set the case of potassium, which has much the same elastic properties as lithium; the θ value for potassium (99) shows no variation as far as the measurements go (Simon and Zeidler, 1926). Further, the θ value from elastic data at low temperatures (Bender, 1937) is 96·5. It seems certain, therefore, that even if a variation in the θ value should occur at lower temperature than those measured, the behaviour will be quite different from that of lithium.

The agreement between Debye theory and experiment becomes surprisingly different, however, if one considers lower temperatures, $T<\theta/12$, i.e. the region where the T^3 law " ought " to hold (*vide* table 7).

Table 7

NaCl			KCl			Li		
$T°$ K.	θ	$\frac{C_v}{T^3}\cdot 10^4$	$T°$ K.	θ	$\frac{C_v}{T^3}\cdot 10^4$	$T°$ K.	θ	$\frac{C_v}{T^3}\cdot 10^4$
20	288	0·388	14	213	0·960	30	356	0·101
15	297	0·356	8	222	0·832	20	340	0·118
10	308	0·334	4	236	0·708	15	328	0·131
			3	227	0·798			

In the above table C_v is, in the case of NaCl and KCl, the molar heat capacity and, in the case of Li, the atomic heat capacity.

It will be seen that the T^3 law is rather elusive. In both lithium and NaCl it seems that the ratio C_v/T^3 is tending to become constant, but there is no sign of a constant region having been reached at the lowest temperature.

The variation for KCl (Keesom and Clark, 1935) is rather peculiar, especially below 4° K., and this may be spurious; Clark (1935) suggests that this " dip " in the value of θ may be due to the desorption of helium adsorbed on to the sylvine —a small quantity of helium had to be admitted to the calorimeter in order to ensure good heat contact. This view is supported by the difference in the θ value at 3° K. and that derived from elastic data ($\theta = 246$). What is clear in any event is, that for such a " standard " crystal as KCl the T^3 region does not start above 4° K., i.e. even at $T = \theta/50$.

Similar results are obtained when one considers silver and zinc (table 8).

Table 8

Silver				Zinc			
$T°$ K.	C_v	$\frac{C_v}{T^3} \cdot 10^4$	θ	$T°$ K.	C_v	$\frac{C_v}{T^3} \cdot 10^4$	θ
1·671	0·0004149	—	173·5	2·0	0·000386	—	212·6
2·037	0·0006537	—	181·8	3·2	0·000935	—	253·8
3·079	0·001664	—	200·9	4·0	0·00155	—	268·5
3·534	0·002253	—	208·6	6·0	0·00476	—	265·0
5·4	0·00634	0·401	226	10·0	0·0387	0·387	229·2
7·0	0·0151	0·439	219	14·0	0·137	0·499	210·2
10·0	0·0475	0·475	214	16·0	0·219	0·534	205·6
14·0	0·1336	0·486	212	12·95*	0·1140	0·511	207
16·0	0·2020	0·493	211	14·45*	0·1645	0·542	204
18·0	0·2898	0·497	210	16·40*	0·233	0·528	206
20·0	0·3995	0·500	209	18·58*	0·344	0·535	205

The above data are due mainly to Keesom and van den Ende (1932) and Keesom and Kok (1932) ; the starred values are taken from a paper by Clusius and Harteck (1928).

It will be seen that there is a small region where C_v/T^3 is nearly constant, followed, however, by a drop as the temperature is decreased, i.e. a rise in the θ value. This rise would be further accentuated if allowance were made for the heat capacity of the free electrons, which becomes important below 10° K. and which is responsible for the final drop in the θ value. One can attempt to separate the two effects, as the electronic specific heat will be proportional to T, but this is not necessary for our purpose. The main points are that the agreement with the Debye theory is by no means good in this region of very low temperature, that there is no clear sign of the true T^3 region, and that there is a spurious T^3 region over a short range of temperatures near 20° K.

There are, however, two cases in which very good agreement is obtained. These are the standard example, CaF_2, and FeS_2; the constancy of C_v/T^3 is practically perfect over a range of θ/T from 12 to 25 in each case. The only point which arises here is that both the θ values, 470 and 645, are lower than the values obtained from elastic constants at room temperature (namely, 499 and 682 respectively), and the discrepancy would be increased if low-temperature values were taken.

The experimental evidence for the existence of a T^3 region in the specific heat is, on the whole, weak (cf., in this connection, Eucken and Werth, 1930); in no case has it been clearly demonstrated that the true T^3 region has been reached; this would involve correspondence between " thermal " and " elastic " values of θ, and there is no case where there is exact agreement. What evidence there is on this point seems to be that the true T^3 region, if it exists at all, will be in the region below 4° K. for most substances.

It is also evident that the good agreement of the Debye theory with experiment does not in general extend to the region of very low temperatures. It is on this point that the lattice theory of specific heats has been able to throw some light.

§ 5. LATTICE THEORY OF SPECIFIC HEATS

The general explanation of the effects discussed in § 4 seems to be that the Debye theory deals with a crystal as if it were an elastic continuum, and that the specific

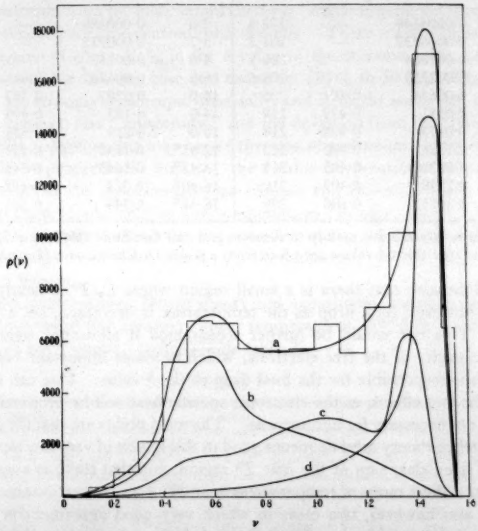


Figure 3 (*a*). Vibrational spectrum of a simple cubic lattice, curve *a*, together with the three branches *b*, *c*, *d* which make up the spectrum (Blackman, 1937 a).
(*Reproduced by permission of the Council of the Royal Society.*)

crystal properties are of much greater significance than had been realised. In particular, the vibrational spectrum of a crystal can be very different from a Debye spectrum. This was first demonstrated in an investigation on the

vibrational spectrum of a two-dimensional lattice (Blackman, 1935 b), and confirmation of this was obtained in a calculation of the spectrum of a simple cubic lattice (Blackman, 1937) of the type discussed by Born and v. Kármán (1912) in their first paper on the specific heats of solids. They worked out the equations for the frequencies of a lattice of the NaCl type, with forces between neighbours and " next neighbours ", and showed that the model represented a fair approximation (cf. also the work of Thirring, above).

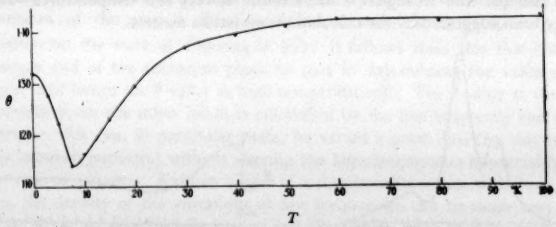


Figure 3 (*b*). θ,T curve for the vibrational spectrum of figure 3 (*a*) (Blackman, 1937 a).
(*Reproduced by permission of the Council of the Royal Society.*)

The spectrum obtained with a particular model is shown in figure 3 (*a*). The variation of the θ value with temperature (when the specific heat is calculated from this spectrum) is shown in figure 3 (*b*). The constants involved in the calculation were chosen, for convenience, in a manner which made the lattice rather anisotropic elastically, more so than is the case for KCl—which is otherwise the nearest approach to the model.

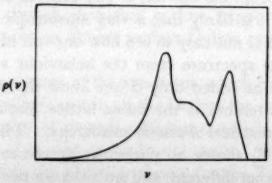


Figure 4. Vibrational spectrum of a body-centred cubic lattice (Fine, 1939).
(*Reproduced from the ' Physical Review ' by permission of the American Physical Society.*)

Calculations on the lines of a Born-v. Kármán model have been made by Fine (1939) for a body-centred lattice with the force constants chosen so as to fit the case of tungsten. The spectrum is shown in figure 4. Though one would have thought such a model a rather poor substitute for a metal, the calculation does show the main features of the specific-heat curve (the increase in θ as the temperature is lowered (cf. table 6)), without, however, giving complete agreement.

A careful investigation of NaCl has recently been made by Kellermann (1940, 41) using ionic and repulsive forces between the particles. The resultant spectrum (figure 5 (*a*)) shows the same general features as that obtained with the Born-v. Kármán model, and the specific heat shows remarkable agreement with experiment (figure 5 (*b*)). The work of Kellermann confirms the view that the deviation of the specific heat of NaCl from a Debye curve below 20° K. is due to the inadequacy of that theory and represents no new phenomenon. It also shows that the true T^3 region is to be found at very low temperatures—as has already been suggested by the calculations on lattice models.

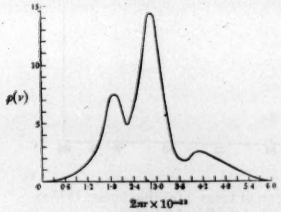


Figure 5 (*a*).

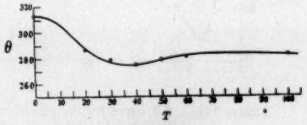


Figure 5 (*b*).

Figure 5 (*a*). Vibrational spectrum of sodium chloride (Kellermann, 1941).
Figure 5 (*b*). θ,T curve for sodium chloride. The smooth curve is constructed from experiment. The points shown are calculated (Kellermann, 1941).
(*Reproduced by permission of the Council of the Royal Society.*)

A calculation of the vibrational spectrum of lithium due to Fuchs (1936 b), is based on the assumption that one can obtain the spectrum from the " continuum " end alone, with an appropriate cut-off, at any rate for such an extremely anisotropic case. Though it is likely that a very anisotropic substance will have an anomalous spectrum, it is not easy to see how one can obtain more than a vague indication of the whole spectrum from the behaviour at very low frequencies. The assumption could be tested only if one knew the whole spectrum; a case which fulfils this requirement is the cubic lattice discussed above (Blackman, 1937), which shows a great deal of elastic anisotropy. The comparison with three Debye spectra (figure 2) shows no particular agreement; though the method used by Fuchs is somewhat different, and smooths out the sharp peaks, the general principle of the method is the same as that which gives the three Debye spectra, and the agreement with the true spectrum cannot be expected to be good.

§ 6. LATTICE THEORY AND DEBYE THEORY

In order to understand the details of the variation of the specific heat with temperature, it would seem to be necessary to calculate the spectrum in each case. Beside being in general extremely tedious, if not impossible, such calculations are hardly likely to throw much light on general questions. For instance, the relation of the Debye theory to the lattice theory is of some interest, especially

as the lattice theory has no alternative " formula " to offer to experimental physicists.

One of the main features of the Debye function is its dependence on one parameter only. Lattice theory suggests that a one-parametric representation cannot be valid in general. To show this one can consider the specific heat in two limiting cases—high temperatures and low temperatures (T^3 region). At high temperatures the specific heat is controlled by the sum of the squares of the frequencies of the normal vibrations of the crystals, i.e. by $\int \rho(\nu)\nu^2 d\nu$ (*cf.* the discussion on the work of Thirring in § 2); it follows from this that the low-frequency end of the spectrum plays no part in determining the value of the integral (and hence the θ value at high temperatures). The θ value at very low temperatures, on the other hand, is controlled by the low-frequency end of the spectrum; this can, in particular cases, be varied a great deal (by varying the forces between particles) without altering the high frequencies appreciably, e.g. in the case of a Born-v. Kármán model of a simple cubic lattice with weak shear forces, the density of the vibrations at low frequencies can be made very large without changing the high-frequency end (Blackman, 1935 b). It can further be seen that in the general case the θ values at low and at high temperatures depend on quite different combinations of the forces between particles, so that complete agreement cannot be expected.

So far only the θ values at the two ends of the specific-heat curve have been considered. The specific heat in the intermediate region is controlled by the spectrum as a whole. The resemblance of the calculated spectra to the Nernst-Lindemann spectrum is rather striking, and, as pointed out by Debye (1912), the two functions give practically the same specific-heat function over a large range of temperatures, though deviating at low temperatures. In particular cases this might account for an almost perfect Debye curve at moderately high temperatures.

A rather interesting property of the spectrum, which has some bearing on the variation of specific heats, is found in all the lattices for which the spectrum has been analysed; the density of normal vibrations, which starts off as a ν^2 law, seems to rise faster than ν^2 as soon as the " continuum " region is passed. This was shown to be a consequence of lattice theory for the Born-v. Kármán models of the simple cubic lattice (Blackman, 1937 b), and it can also be seen to be a general property of lattices in which short-range forces are employed. No general proof has so far been obtained, and it is by no means certain that the law is a general one; the same effect does, however, occur in ionic crystals as well, as can be seen by examining the work of Kellermann (1940). The linear (ionic) chain also shows this property (Broch, 1937).

This property of the spectrum means that the value of C_v/T^3, which should be constant at " sufficiently " low temperatures, will show an increase as the temperature is raised; in terms of Debye theory the θ value will drop, the magnitude of the " drop " depending, however, very much on particular properties

of the spectrum. It is this effect which moves the true T^3 region to such low temperatures—in the case of the lattices investigated to a temperature $T<\theta/50$, possibly even to $T<\theta/100$ (Blackman, 1937, p. 430). This means, for all practical purposes, no T^3 region at all, as measurements at helium temperatures are a rarity.

In view of all the complications which occur in the lattice theory of specific heats, the Debye theory is remarkable, not in that it fails in some cases rather badly, but for the fact that it succeeds so well for most crystals. One is less surprised at deviations of, say, 30 % in the θ value (e.g. lithium) than at a deviation of less than 10 % (e.g. NaCl, KCl and Ag).

§ 7. MISCELLANEOUS APPLICATIONS OF LATTICE THEORY

The deviation of the specific heat from a Debye curve for some substances, e.g. lithium, led to the hypothesis that there is an extra term in the specific heat due to an " electronic excitation " (Simon, 1926). The case of lithium has aroused particular interest as the variation of the expansion coefficient with temperature —which in most crystals follows the same course as the specific heat—showed a different behaviour ($\theta=525$, instead of 330–430). As long as there is no actual maximum in the specific-heat curve, there is no reason to suppose that the lattice theory alone will not account for the variation of specific heat; in the case of lithium the recently calculated (Fuchs, 1936) and measured (Bender, 1937) elastic constants give values of θ—namely, 354 and 333 respectively—which agree quite well with that obtained from the specific heats at low temperatures (328). This forms conclusive evidence that, in the case of lithium, a value of 525 is incorrect, and hence the assumption of an extra specific-heat term is unnecessary. The evidence for sodium and potassium also indicates that the vibrational specific heat alone will be sufficient to account for the observed specific heats.

It has been the practice in early accounts of the theory of specific heats to compare the high-frequency end of the spectrum (or the θ value) with that obtained from other data, e.g. from *Reststrahlen* in the case of ionic crystals, from the variation of the expansion coefficient and from the variation of resistance with temperature. All these methods yield frequencies (or θ values) which agree more or less with that obtained from a Debye curve, but it is difficult to see what the comparison actually means. The spectrum of NaCl, for example, extends well above the frequency at which the main absorption of infra-red radiation takes place, and also well above the Debye frequency ($\nu_D=k\theta/h$). The danger of comparing values from different phenomena is well illustrated by the case of lithium, discussed above. Until the theories of the variation of the expansion coefficient and of the variation of resistance with temperature have been more carefully analysed, it is as well not to draw conclusions from apparent discrepancies.

§8. GENERAL CONCLUSIONS

The main feature in the recent developments of the theory of the specific heat of solids has been the emergence of the vibrational spectrum, which shows a different character from that assumed by Debye; this represents a vindication of the method adopted originally by Born and v. Kármán, which method, however, was not carried out completely at the time. The properties of the vibrational spectrum explain the deviations from Debye theory observed at low temperature for the (vibrational) specific heat, and enable one to clear up points such as the existence of T^3 regions which are not true T^3 regions. These latter regions occur in general at much lower temperatures than one would expect on Debye theory, in many cases at such low temperatures that they are not likely to be of practical interest.

GENERAL REFERENCES

BORN, M. and GOEPPERT-MAYER, M., 1933. *Handbuch der Physik*, **24** (2).
EUCKEN, A., 1929. *Handbuch der Experimentalphysik*, **8** (1).
FOWLER, R. H. and GUGGENHEIM, E. A., 1939. *Statistical Thermodynamics* (Cambridge).
SCHRÖDINGER, E., 1926. *Handbuch der Physik*, **10**.

OTHER REFERENCES

BEHN, U., 1898. *Wied. Ann.* **66**, 237.
BENDER, O., 1937. *Diss.* (Marburg).
BLACKMAN, M., 1935 a. *Proc. Roy. Soc.* A, **148**, 365.
BLACKMAN, M., 1935 b. *Proc. Roy. Soc.* A, **148**, 384.
BLACKMAN, M., 1935 c. *Proc. Roy. Soc.* A, **149**, 117.
BLACKMAN, M., 1935 d. *Proc. Roy. Soc.* A, **149**, 126.
BLACKMAN, M., 1937 a. *Proc. Roy. Soc.* A, **159**, 416.
BLACKMAN, M., 1937 b. *Proc. Camb. Phil. Soc.* **33**, 94.
BORN, M. and BRODY, E., 1921. *Z. Phys.* **6**, 132.
BORN, M., 1923. *Atomtheorie des festen Zustandes* (Leipzig).
BORN, M. and v. KÁRMÁN, TH., 1912. *Phys. Z.* **13**, 297.
BORN, M. and v. KÁRMÁN, TH., 1913. *Phys. Z.* **14**, 15.
BROCH, E., 1937. *Proc. Camb. Phil. Soc.* **33**, 485.
CLARK, C. W., 1935. *Diss.* (Leiden).
CLUSIUS, K. and HARTECK, P., 1928. *Z. phys. Chem.* **134**, 243.
CLUSIUS, K. and VAUGHAN, J. V., 1930. *J. Amer. Chem. Soc.* **52**, 4680.
DEBYE, P., 1912. *Ann. Phys., Lpz.*, **39**, 789.
DURAND, M. A., 1936. *Phys. Rev.* **50**, 453.
EINSTEIN, A., 1907. *Ann. Phys., Lpz.*, **22**, 180 and 800.
EINSTEIN, A., 1911. *Ann. Phys., Lpz.*, **34**, 170 and 590.
EUCKEN, A. and SCHWERS, F., 1913. *Verh. dtsch. phys. Ges.* **15**, 580.
EUCKEN, A. and WERTH, H., 1930. *Z. anorg. Chem.* **188**, 152.
EUCKEN, A., CLUSIUS, K. and WOITENECK, H., 1931. *Z. anorg. Chem.* **203**, 1931.
FINE, P. C., 1939. *Phys. Rev.* **56**, 355.
FÖRSTERLING, K., 1920 a. *Ann. Phys., Lpz.*, **61**, 549.
FÖRSTERLING, K., 1920 b. *Z. Phys.* **3**, 9.
FÖRSTERLING, K., 1920 c. *Z. Phys.* **8**, 251.
FUCHS, K., 1936 a. *Proc. Roy. Soc.* A, **153**, 622.
FUCHS, K., 1936 b. *Proc. Roy. Soc.* A, **157**, 444.
GRÜNEISEN, E. and GOENS, E., 1924. *Z. Phys.* **26**, 235 and 250.

HONNEFELDER, K., 1933. *Z. phys. Chem.* B, **21**, 53.
HOPF, J. and LECHNER, G., 1914. *Verh. dtsch. phys. Ges.* **16**, 643.
KEESOM, W. H., 1934. *Phys. Z.* **35**, 939.
KEESOM, W. H. and CLARK, C. W., 1935. *Physica*, **2**, 198.
KEESOM, W. H. and VAN DEN ENDE, J., 1932. *Proc. Acad. Sci. Amst.* **35**, 143.
KEESOM, W. H. and KOK, J. A., 1932. *Proc. Acad. Sci. Amst.* **35**, 301.
KELLERMANN, E. W., 1940. *Philos. Trans.* A, **238**, 513.
KELLERMANN, E. W., 1941. *Proc. Roy. Soc.* A, **178**, 17.
LANGE, F., 1927. *Z. phys. Chem.* **110**, 351.
LANGE, F. and SIMON, F., 1928. *Z. phys. Chem.* **134**, 374.
LEWIS, G. N. and GIBSON, G. E., 1917. *J. Amer. Chem. Soc.* **39**, 2565.
NERNST, W. and LINDEMANN, F., 1911. *Z. Elektrochem.* **17**, 817.
SIMON, F., 1926. *Ber. Berl. Akad.* 477.
SIMON, F. and SWAIN, R. C., 1935. *Z. phys. Chem.* B, **28**, 189.
SIMON, F. and ZEIDLER, W., 1926. *Z. phys. Chem.* **123**, 383.
SOMMERFELD, A. and BETHE, H., 1933. *Handbuch der Physik*, **24** (2).
SOUTHARD, J. C. and NELSON, R. A., 1933. *J. Amer. Chem. Soc.* **55**, 4865.
THIRRING, H., 1913. *Phys. Z.* **14**, 867.
THIRRING, H., 1914. *Phys. Z.* **15**, 127 and 180.
ZWIKKER, C., 1928. *Z. Phys.* **52**, 668.

HIGH-SPEED CENTRIFUGING

By J. W. BEAMS,
University of Virginia, Charlottesville, Virginia

§ 1. INTRODUCTION

In addition to its long use as a means of separating or purifying materials of different density, the centrifuge in recent years has become a powerful analytical tool for measuring sizes, weights and shapes of particles. Since, in general, the latter use is more exacting on the apparatus than the former, careful studies, both theoretical and experimental, of the centrifuging process have been made; these have resulted in several fundamental advances. The principal theoretical advances have consisted in the derivation of the relations between the amount of sedimentation in dilute idealized solutions and the physical and chemical properties of the solvent and solute, together with the derivation of the relations for the separation of gases. Experimentally, the advances have been made mainly by increasing the effective centrifugal field and in preventing re-mixing of the material during the centrifuging process.

§ 2. THEORY OF SEDIMENTATION

The solution of the differential equation for sedimentation in an ideal dilute solution in a uniform gravitational field was first worked out by Mason and Weaver (1924). This theory, while not strictly applicable to the sedimentation in a centrifuge, where the field varies as the radius, may be used to give approximate results in a centrifuge where the radius of the centrifuge is large and the length of the sedimenting column small. The extension of this theory by Weaver (1926) is of considerable use in calculating the time required for equilibrium between sedimentation and diffusion to be established in a centrifuge.

The general differential equation for the sedimentation of uncharged particles or molecules in an ideal, dilute, incompressible solution enclosed in a sector-shaped centrifuge cell rotating with an angular velocity ω is, according to Lamm (1929a, 29b, 37),

$$\frac{1}{r}\frac{\partial}{\partial r}\left\{\left(D\frac{\partial c}{\partial r}-\omega^2 rsc\right)r\right\}=\frac{\partial c}{\partial t}, \qquad \ldots\ldots(1)$$

where r is the distance of the material from the axis of rotation, c is the concentration, t the time, D the diffusion constant, and s the sedimentation constant or the velocity of settling of the solute (or particles) in unit (gravitational) field. Faxen (1929, 36) and Soyten (1936, 37) have each given approximate solutions to equation (1) which, although indicating the possible error introduced into certain

measurements, are usually not suited for numerical calculations. More recently Archibald (1938 a, 38 b) has obtained a complete solution to equation (1) and has been able to put it into a usable form. He first (1938 a) treated the case of a sector-shaped centrifuge cell extending to the axis of rotation, and afterwards (1938 b) worked out the case where the sector cell was bounded by any two coaxial cylindrical surfaces. Fortunately, in most centrifuging experiments it is possible to utilize the simple solutions of equation (1), derived many years ago by Svedberg (1928–39) * for special cases, rather than the more elaborate and detailed general solutions of Archibald. Svedberg has shown that if the centrifuging process is continued until equilibrium between sedimentation and diffusion is established, i.e., $\partial c/\partial t = 0$ in equation (1), and the change in free energy throughout the cell vanishes, then

$$M_e = \frac{2RT \log_e (c_1/c_2)}{(1 - V\rho)\omega^2(r_1^2 - r_2^2)}, \qquad \ldots\ldots(2)$$

$$v_e = \frac{2RT \log_e (c_1/c_2)}{N(\rho_p - \rho_d)\omega^2(r_1^2 - r_2^2)}, \qquad \ldots\ldots(3)$$

where M_e is the molecular or particle weight, v_e is the volume of the particle or molecule, c_1 and c_2 are the concentrations at distances r_1 and r_2 from the axis of rotation respectively, V is the partial specific volume of the substance, ρ, ρ_p and ρ_d are the densities of the solution, particle and solvent respectively, R is the gas constant, T is the absolute temperature, and N is the Avogadro number.

Since the quantities entering into equation (2) are measured for equilibrium conditions in a dilute ideal solution, the M_e obtained is independent of the shape of the molecule (assuming that any small orientation of non-spherical particles by the field does not affect diffusion). In practice V is usually measured outside the centrifuge and consequently does not take account of hydration or solvation effects. Kraemer (1941), and his collaborators † have shown how these solvation effects can be estimated, and conclude that usually they are small, though not in all cases. If the solution contains more than one solute, the concentration will be given, by equation (2), as a function of the radius for each, independently of the others. In instances where the concentrations in the centrifuge cell are measured by optical or other methods which do not distinguish directly between the different molecular species, the mixture manifests itself by a drift in M_e for different respective c and r values. It is possible not only to determine the molecular weights of the different homogeneous molecular species in a solution, but also to get the distribution of particle or molecular weights. However, in cases where the molecular weights are not too small and the molecular shapes are not thread-like, it is usually more convenient to use the rate-of-sedimentation method, to be described later, for this latter case. Equations (1), (2) and (3) were derived on the assumption that the particles or molecules were uncharged.

* See also Svedberg and Pedersen, 1940.

† See Kraemer and Lansing, 1934–36; Svedberg and Pedersen, 1940.

Frequently in cases where the solutions are not at the isoelectric point, or when the particles are charged, the electrostatic effect of the ions changes the sedimentation. If the molecules or particles are large, these charges can be repressed by the addition of a low molecular weight electrolyte such as NaCl (Svedberg, 1928–39). For an electrolyte which dissociates into m n-valent cations and n m-valent anions, where the numbers n and m contain no whole-number factor, Pedersen (Svedberg and Pedersen, 1940) has shown that

$$M_e = \frac{2(m+n)RT \log_e (c_1 f_1^{\pm}/c_2 f_2^{\pm})}{(1-V\rho)\omega^2(r_1^2 - r_2^2)}, \qquad \ldots\ldots(4)$$

where $f_1^{\pm}$ and $f_2^{\pm}$ are the " activity coefficients ".

Although a minimum of measurements is required to obtain M_e by equations (2) and (4) and the equations hold with high approximation, it is often necessary to centrifuge for prohibitively long times before equilibrium is established. This is especially true in the case of the large molecular weight substances, which may decompose with time. Archibald (1938 a, 38 b) has given a method of calculation for the equilibrium time, but the criterion developed by Weaver (1926) for a uniform field can usually be applied to the centrifuge cell to give the approximation desired. Weaver showed that, in a uniform gravitational field, equilibrium is established in a time t_{max} required for the particles to settle through twice the vertical length of the sedimenting column. Applied to the centrifuge (see Svedberg and Pedersen, 1940) this gives

$$t_{max} = \frac{(r_1 - r_2)^2}{1800 D_{20} \log_e (c_1/c_2)} \text{ hours}, \qquad \ldots\ldots(5)$$

where D_{20} is the diffusion constant at 20° C. Thus it is seen that, for some of the biologically important substances whose molecular weights are of the order of 10^5 to 10^6, t_{max} is almost prohibitively long.

Fortunately, a second method of determining molecular weights and sizes, also due to Svedberg (1928–39), and known as the rate-of-sedimentation method, supplements the equilibrium method discussed above, i.e., it is especially applicable to the case of large molecular weight compounds. If the solution is ideal and dilute and the particles or molecules are uncharged, also if the reflection of the molecules or particles from the ends of the sedimenting column can be neglected, then by equating the sedimentation per mole to the frictional force per mole,

$$M_s(1-V\rho)\omega^2 r = f.\, dr/dt, \qquad \ldots\ldots(6)$$

where $f = RT/D$ from the diffusion theory, and

$$M_s = \frac{RT}{D(1-V\rho)} \frac{dr/dt}{\omega^2 r} = \frac{RT}{D(1-V\rho)} s, \qquad \ldots\ldots(7)$$

where s is the sedimentation constant or velocity of sedimentation in a unit field. Since each homogeneous molecular species has its characteristic s, it is convenient

to reduce sedimentation constants to a standard solution at a standard temperature (20° c.) by the equation

$$\frac{s_t}{s_{20}} = \frac{\eta_{20}}{\eta_t} \frac{1 - V_s\rho_t}{1 - V_{20}\rho_{20}}, \qquad \ldots\ldots(8)$$

where η_t, V_s and η_{20}, V_{20} are the viscosities of the medium and the partial specific volumes of the solute in the solution at $t°$ and 20° c., and ρ_t and ρ_{20} are the densities of the medium at $t°$ c. and of the reference medium at 20° c.

If the particles are spherical and Stokes's law is assumed, then the radius a of the particle or molecules is given by

$$a = \left[\frac{9}{2} \frac{\eta \log_e(r_2/r_1)}{(\rho_p - \rho_d)\omega^2 t}\right]^{\frac{1}{2}}, \qquad \ldots\ldots(9)$$

where η is the coefficient of viscosity and t is the time of centrifuging. If the particles are non-spherical, corrections must be made to equation (9). It will be observed that in order to obtain M_s in equation (7), one must measure the rate of settling of the material dr/dt in the centrifuge cell and the diffusion constant D in addition to T, V, ρ and ω. In most cases the rate-of-sedimentation method is useful only when the centrifugal force is high enough or the molecular weight large enough to form an easily distinguishable boundary between the clear solvent and the sedimenting substance. Hence, by observing the rate at which this sedimenting boundary moves in the centrifugal cell, the sedimentation constant can be calculated. As the boundary moves outward, a blurring due to diffusion takes place. Svedberg (1928–39) has shown that the concentration c_x at a distance x (positive toward the axis) from the sharp sedimenting boundary (which would have been formed if diffusion were absent) is given by

$$c_x = \left(1 - 2\pi^{-\frac{1}{2}} \int_0^y e^{-y^2} dy\right) c_i/2, \qquad \ldots\ldots(10)$$

where $y = x/2(Dt)^{\frac{1}{2}}$ and $c_i/2$ is the concentration at $x = 0$ after the time of centrifuging t. Hence, from the measurements of the concentration in the centrifuge cell as a function of the distance from the axis of rotation, both s and D can be obtained. However, it has been found by experience that it is often advisable to measure D by a separate experiment. Methods of measuring the partial specific volume while the centrifuging is taking place have not proved satisfactory, so it is necessary to obtain V by experiments done outside the centrifuge. Consequently, as in the equilibrium method of equation (2), the effect of hydration is difficult to estimate. As a matter of fact, the uncertainties due to hydration effects are more serious in the case of the rate-of-sedimentation method than in the sedimentation equilibrium method. The effects of solvation on the M_s values obtained by equation (7) have been critically discussed by Kraemer and his colleagues.* The shape of the particle or molecule also affects the sedimentation constant s and the values of M_s obtained by equation (7). Experiment shows that in the case of those substances whereboth M_s and M_s values have

* See Kraemer and Lansing, 1934–36; Svedberg and Pedersen, 1940.

been measured, they often differ by appreciable amounts. In those substances where this difference occurs, the evidence points either to the existence of hydration or non-spherical molecules or both. In those cases where the possible effects of hydration are known to be small, Svedberg has indicated how the shape of the molecules can be deduced from centrifuging data (Svedberg, 1928–39 ; Svedberg and Pedersen, 1940).

When the solution contains more than one sedimenting substance, each homogeneous molecular species forms its characteristic sedimenting boundary, so that the rate-of-sedimentation method analyses the solution by giving the molecular weights of all the substances it contains. When the solution contains a large number of molecular species or a distribution of molecular sizes or weights, the rate-of-sedimentation method gives the molecular distribution.

As pointed out before, the theory discussed above holds strictly only for ideal dilute solutions. In practice these conditions are sometimes not fulfilled with sufficiently close approximations, and it is necessary to estimate the necessary corrections. Obviously there is need for a general theory of sedimentation in liquids in a centrifuge which is applicable to all cases, i.e., to concentrated non-ideal solutions. Such a theory, if obtained, might be difficult to use in practice and would probably depend to a certain extent upon the structure of the liquid ; but it would undoubtedly give valuable information as to the magnitude of the corrections necessary to the present theory.

§3. THEORY FOR THE SEPARATION OF GASES AND VAPOURS IN A CENTRIFUGE

The separation of gases in a centrifuge has been treated theoretically for special cases by Chapman (1919), Lindemann and Aston (1919), Joly and Pool (1920), Joly (1921) and Mulliken (1922 a, 22 b). For a gas that follows the ideal gas laws

$$\rho_r = \rho_0 \exp\,(M\omega^2 r^2/2RT), \qquad \ldots\ldots(11)$$

where ρ_0 is the density of the gas near the axis and ρ_r is the density at a distance r from the axis of rotation. If the gas or vapour contains two isotopes of masses M_1 and M_2 and equilibrium is established in the centrifuge, Lindemann and Aston showed that

$$S = K_0/K = \exp\,(M_2 - M_1)\omega^2 r^2/2RT, \qquad \ldots\ldots(12)$$

where K_0 is the ratio of the quantities of the light and heavy isotopes at the axis and K is the same ratio at a distance r from the axis. They defined S as the separation factor.

This theory has been extended by Mulliken to include the separation of the isotopes in substances containing several isotopes. Also, he has shown that the separation factor for any two isotopes of an element in a centrifuge is independent of the state of combination of the element. Chapman treated the problem as one of separation by pressure diffusion, and arrived at essentially the same results

as Lindemann and Aston and Mulliken. It will be observed that according to equation (12) the separation factor increases very rapidly with increasing mass-differences and peripheral speeds. Recently the separation of the isotopes in a centrifuge when equilibrium does not exist has been investigated by Beams (1937, 38b, 40), Beams and Skarstrom (1939), Humphreys (1939), Urey (1940) and Wilson (1940). They of course find that equation (12) must be modified when equilibrium does not exist, but the corrections depend upon the particular method of centrifuging. In addition to the separation of isotopes, equation (12) can be used to compute the separation of any two gases which approximately obey the ideal gas laws.

§4. EXPERIMENTAL TECHNIQUE

The modern technique of high-speed centrifuging has been developed almost completely in the last decade and a half. Beginning in 1924, Svedberg first succeeded in obtaining convection-free sedimentation in centrifugal fields as high as 5000 g. The results obtained with this machine proved to be so interesting and important that Svedberg and his collaborators were led to undertake an extensive development of centrifuging technique. This work has resulted in the development of their two general types of centrifuges which have played an indispensable rôle in their classical experiments on the determination of molecular weights and sizes. Both types of Svedberg ultracentrifuges were designed primarily for analytical work, i.e., the determination of molecular weights and sizes.

The first type develops comparatively low centrifugal fields (500 to 15,000 g). It is driven by an electrical motor and supported in ball bearings. The rotor carries a cell which contains the material to be centrifuged. The cell is usually sector-shaped and has glass or quartz windows through which the concentration of the material being centrifuged at various radial distances can be determined by optical means while the rotor is spinning. The rotor spins in hydrogen at atmospheric pressure and is surrounded by a water-cooled casing. This machine has been used principally for sedimentation-equilibrium measurements (equation (2)). The second type of Svedberg ultracentrifuge is driven by oil turbines and is used to produce fields in the range between 15,000 and 750,000 g. The rotor of this machine is made of a special alloy steel and is designed to give a maximum centrifugal field inside the centrifuge cell without explosion. The cell which carries the material to be centrifuged is sector-shaped and is enclosed by crystal quartz windows for observing optically the concentration of the sedimenting material while the rotor is spinning. The rotor is surrounded by a heavy steel case through which hydrogen at about 20 mm. of mercury pressure is continuously circulated. The purpose of the hydrogen is to conduct the heat generated by the friction in the bearings, by the gaseous friction on the rotor, and by the oil impinging on the turbine, to the casing. By proper adjustment this makes it possible to obtain a uniform temperature throughout the centrifuge

cell. The rotor and its contents are carefully balanced, both statically and dynamically, before running the machine. This ultracentrifuge has been used primarily for rate-of-sedimentation measurements (equation (7)). For detailed descriptions of these machines, reference should be made to the papers by Svedberg (1928–39) and his collaborators (see Svedberg and Pedersen, 1940).

§ 5. AIR-DRIVEN CENTRIFUGES

There are several different kinds of high-speed air-driven centrifuges, but they can be classified somewhat arbitrarily into two general types. In the first type the centrifuge rotor spins in air at atmospheric pressure, while in the second type the centrifuge rotor is enclosed in an air-tight chamber which may be evacuated. The first type is primarily suited to centrifuging small amounts of material and where accurate temperature control is not essential. The second type has a wider use and can be applied to practically every type of high-speed centrifuging problem. Also, it is provided with precise temperature control. With either type high rotational speeds are obtained.

A significant advance in the first type of air-driven centrifuge was made in 1925 by Henriot and Huguenard (1925, 27), when they succeeded in supporting and spinning small rotors at very high speeds (11·7 mm. rotor to 11,000 r.p.s.) on properly directed jets of air. The rotor was cone-shaped and spun a fraction of a millimetre above the surface of a cone-shaped stator into which it fitted. This design of Henriot and Huguenard has been modified, improved, and stabilized by Beams (1930 a–40), by Garman (1933), by Girard and Chukri (1933) and by McBain (1939). Thus, at the present time, the centrifuges of this type are remarkably stable and are also simple and easy to construct (Beams, 1937, 38 b, 40). Since this type of centrifuge is driven and supported by expanding jets of air, the lower part of the rotor is cooled, while the top part and the periphery are heated by air friction. Consequently, unless special care is taken, temperature gradients large enough to produce re-mixing or convection of the material being centrifuged may exist in the centrifuge cell. Incidentally, convection due to differences in density resulting from temperature variations are greatly magnified in a centrifuge because of the large centrifugal field (Beams, 1937, 38 b, 40).

This type of centrifuge has been used widely for centrifuging small biological tissues or cells which have such minute volumes that the temperature gradients in the centrifuge have no appreciable effect upon the sedimentation inside the biological cell. The relative movements of the various components of the biological cell in the centrifugal field may be determined by microscopic examination after the centrifuge is stopped, or under comparatively low-power magnification while the rotor is spinning (Beams (H. W.) and King, 1934–38 ; Harvey, 1932–38 ; Pickels, 1936, 38). Also, many workers have used this type of centrifuge for the sedimentation of small particles and molecules ; it has also been successfully applied to analytical work by Beams, Weed and Pickels (1933), by McIntosh and Selbie (1937), by Beams (1937, 38 b, 40), by Bawden and

Pirie (1938), by Elford (1938) and by McBain (1939). McBain (1939) and McBain and Leyda (1938 a, 38 b) especially have found this type of centrifuge useful in their molecular-weight studies. Incidentally, these air-supported air-driven turbines have provided means of spinning small mirrors to very high rotational speeds (Henriot and Huguenard, 1925, 27; Beams, 1930 a–40; Chiles, 1937), and have been used to study the light emitted in the various stages of condensed electrical discharges, to demonstrate the existence of the velocity of light (rough measurements), and to produce short flashes of light. Unfortunately, as the size of the mirror is increased, the air friction becomes so great that it is difficult to obtain high rotational speeds with this type of turbine drive. For the rotor shapes used for the above types of centrifuges, the air friction increases very rapidly with increasing radius of rotor for a given peripheral velocity. In practice this limits the diameter of the rotors, which can be spun to high speeds, to an inch or two unless, of course, large high-pressure air supplies are utilized. As a result, in addition to troublesome temperature variations in the rotor noted above, the centrifugal field must vary rapidly along the small radius and thus reduces the precision of the experiments. If hydrogen is used for the driving and supporting gas instead of air, higher peripheral velocities are obtained. For example, a 9 mm. rotor driven and supported by air attained about 12,500 r.p.s., while the same rotor spun by hydrogen at the same pressure as the air attained a rotational speed of about 22,000 r.p.s. (Beams, Weed and Pickels, 1933; Beams, 1937, 38 b, 40).

§ 6. VACUUM-TYPE AIR-DRIVEN CENTRIFUGE

The troublesome effects of air or gaseous friction on high-speed rotors led to the development of the air-driven vacuum-type ultracentrifuge (Beams and Pickels, 1935; Pickels and Beams, 1935; Beams, 1937, 38 b, 40; Beams, Linke and Sommer, 1938). The various designs of these machines differ somewhat, but essentially they consist of a large (centrifuge) rotor situated inside a vacuum-tight chamber, a smaller, air-driven, air-supported turbine located above or below the chamber, and a small-diameter flexible shaft which fastens them together and is coaxial with their common vertical axis of rotation. The flexible shaft passes through a vacuum-tight oil gland which seals the vacuum chamber. This oil gland has comparatively little friction since the shafts are small, so that a properly designed air turbine will reach a rotational speed of the same order of magnitude with the attached rotor inside the evacuated chamber as outside it. The vacuum attainable in the chamber is limited only by the vapour pressure of the oil, which may be very low (10^{-6} mm. for some oils used). The flexible shaft allows the large rotor to "seek its own axis of rotation" and permits extremely steady rotation, i.e., the rotor is self-balancing. With these machines the only factor which limits the rotational speed attainable is the mechanical strength of the large rotor. This, coupled with the fact that the rotor spins in a vacuum where no appreciable heat which might cause re-mixing, can be generated

makes this vacuum-type machine an ideal high-speed centrifuge for most purposes. The vacuum-tight glands may be thermostated in such a way that material can be forced into and out of the spinning rotor through hollow shafts without appreciably changing its temperature. In addition to rotors whose diameters are greater than their length, the same technique has been used to spin long tubular rotors (Beams, 1930 a–40). In these tubular vacuum-type air-driven centrifuges, the material to be centrifuged (liquid, gaseous, or vapour) enters through a hollow shaft at the top of the vertical spinning tube and passes out at the lower end in a heavy and a light fraction through a shaft with two partitions, or through two coaxial hollow tubes which serve as the shaft. A number of methods of determining the rotational speed of these high-speed centrifuges, as well as means for holding their speed constant, have been devised, but usually some form of the stroboscopic method is used (Beams, 1930 a, 30 b, 33 a, 33 b, 35, 38 a, 39 ; Davis, 1936 ; Kahler, 1938 ; Pickels, 1936, 38 ; Shapiro and Butt, 1937 ; Snoddy and Beams, 1937).

The vacuum-type air-driven ultracentrifuge has been used extensively as an analytical tool for the determination of molecular weights and sizes. It is applicable both to the sedimentation-equilibrium method (equation (2)) and the rate-of-sedimentation method (equation (7)). The cell containing the material to be centrifuged is sector-shaped * and enclosed by crystal quartz windows for observing the sedimentation. The rotor is designed to give a maximum resolving power throughout the cell, where resolving power is qualitatively defined as the ability to distinguish between two homogeneous molecular species with slightly different sedimentation constants or molecular weights. When the centrifuge is used for rate-of-sedimentation measurements, the resolving power is approximately proportional to the product of the centrifugal force and the length of the sedimenting column, i.e., to $\omega^2 x h$, where x is the distance of the sedimenting boundary from the axis of rotation, and h the length of the sedimenting column (Svedberg, 1928–39 ; Svedberg and Pedersen, 1940). It has been shown both by theory and experiment that the resolving power is highest for values of x between 5 and 10 cm. It will be observed that the maximum value of $\omega^2 x h$ for rotors of a given shape is roughly proportional to the square of the peripheral velocity, which in turn is proportional to the strength of the rotor. It might be assumed from this that the maximum resolving power had been reached for a given rotor material. However, the resolving power has recently been considerably increased for the case of large molecular weight compounds by effectively increasing the length of the sedimenting column (Beams, 1930 a–40). Many interesting bio-

* Svedberg (1928–39 ; see also Svedberg and Pedersen, 1940) has shown that for rate-of-sedimentation measurements the centrifuge cell should be sector-shaped. However, for sedimentation-equilibrium measurements this is not so important. As a matter of fact, it has been shown that the time required for equilibrium to be established is longer in the case of sector-shaped cells than for some other shapes.

logical substances have sedimentation constants so large that they settle through an appreciable length of the sedimenting column in a convection-free ultracentrifuge before it can reach full speed. Artificial convection can be introduced to prevent this, but unless special care is taken, undesirable " false " boundaries may develop. Consequently the full resolving power of the centrifuge cannot be used. In the new high-resolving power vacuum-type ultracentrifuge (Beams, 1930 a–40), the solvent is made to flow from the periphery toward the axis at the same rate as the material settles out toward the periphery. This allows sedimenting boundaries to be observed for long periods of time. Hence sedimenting boundaries of homogeneous molecular species with nearly the same sedimentation constants can be observed to separate.

Various optical schemes have been used for measuring the density of the sedimenting material as a function of the distance from the axis of rotation while the centrifuge is spinning (Svedberg, 1928–39 ; Svedberg and Pedersen, 1940). Generally they depend upon the light absorption of the material or upon its refractive index. The light-absorption methods depend upon the fact that in some cases it is possible to find a convenient absorption band of the sedimenting material which is not absorbed by the solvent. By measuring the light absorption in the cell and knowing the relation between this and the amount of absorbing material, the density is determined. In practice it is often necessary to use ultra-violet light in order to find a satisfactory absorption band, and this requires quartz-fluorite lenses (which should have long focal lengths if precision is desired). In some cases the absorption method yields inaccurate results because of Tyndall scattering (Brohult, 1940 ; Svedberg and Pedersen, 1940).

The refractive-index methods are applicable to any transparent solution. Probably the most precise index-of-refraction method is that developed by Lamm (1929 a–37). This consists in photographing a scale through the transparent centrifuge while the centrifuge is spinning. If the scale or its real image is placed just in front of the cell, the image of the scale formed by a long-focus lens on a photographic plate or in the field of view of a telescope is distorted in those regions where the rays of light pass through gradients in refractive index (density gradients) in the centrifuge cell. From these distortions of the scale, the density gradients in the centrifuge cell can be calculated. The Schlieren method is the most direct and convenient index-of-refraction method. It has taken several forms and has been made automatic by Philpot (1938), Andersson (1939) and others. Other refractive-index methods, such as the interferometer method, have not as yet been extensively used. There is a need for a convenient precise optical method, which is inexpensive, for observing sedimentation in a centrifuge cell.

In some cases it is impossible or impracticable to measure the concentration in the centrifuge cell by optical means. In such instances it is necessary to measure the concentration of the material taken from different distances from the axis of rotation by the ordinary methods of analytical chemistry. Different

methods have been devised for collecting these samples without re-mixing (Svedberg, 1928–39; Beams, 1937, 38 b, 40; McBain, 1939). Tiselius, Pedersen and Svedberg (1937), and others have provided the sector-shaped centrifuge cell with a partition perpendicular to the radius which divides it into approximately equal parts. Usually this partition consists of a thin bakelite plate perforated with a large number of small holes and covered with special filter paper. Such a partition does not disturb sedimentation during the centrifuging process, but prevents re-mixing while the centrifuge is decelerated. If Δ is the total change in amount of the material above or below the partition, x the distance from the axis of rotation of the partition, q the area of the partition, ω the angular velocity, c_0 the concentration of the material at the beginning, and t the time of centrifuging, the sedimentation constant s is given by

$$s = -\frac{1}{2\omega^2 t}\log_e\left(1 - \frac{2\Delta}{q \times c_0}\right). \qquad \ldots\ldots(13)$$

It is difficult to calculate ω and t for the acceleration and deceleration period in the arrangement described above, and it is not as precise as the optical methods. However, a vacuum-type ultracentrifuge rotor has recently been designed in which the sector-shaped cell is filled after the rotor reaches full speed (Beams, 1937, 38 b, 40; Stauffacher, Beams and Snoddy, 1941). When the desired centrifuging time t at a constant ω has elapsed, the cell is effectively divided into two approximately equal parts by a displacement method and the centrifuge then brought to rest. From a chemical analysis of one of the two fractions and the original uncentrifuged sample, the sedimentation constant is determined from equation (13). In order to facilitate the chemical analysis by providing larger specimens for analysis, a large number of identical sector-shaped cells instead of one are constructed in the same centrifuge. This method gives good precision, especially in the case of homogeneous molecular species.

In addition to its use as an analytical tool, the vacuum-type centrifuge has made possible the purification of materials in appreciable quantities (Bauer and Pickels, 1936–40; Wyckoff, 1937 a–38 b). When used chiefly for concentrating materials, it is sometimes referred to as the quantity-type vacuum-type ultracentrifuge (Wyckoff, 1937 a–38 b; Beams, Linke and Sommer, 1938). Many of the most interesting substances in biology and medicine often occur naturally in extreme dilution. They are difficult to concentrate by chemical means because in general they are deactivated by high temperature or large changes in the pH values. Fortunately, high centrifugal fields do not deactivate them, so that the high-speed centrifuge may be used for their purification. The rotor in the quantity-vacuum-type centrifuge is made from a large block of metal in which are holes bored at an angle to the vertical. These holes carry Lusteroid (plastic) test tubes which, in turn, hold the material to be centrifuged. One or more vacuum-tight caps seal the centrifuge (Bauer and Pickels, 1936–40; Beams, Linke and Sommer, 1938). This construction permits the use of high centrifugal fields

on comparatively large quantities of material under conditions ideal for concentration. The smaller the angle made by the test tubes with the vertical, the greater is the efficiency down to angles of about ten degrees (Masket, 1940; Wyckoff, 1937 a–38 b).

The above quantity-type centrifuge, while very useful for centrifuging volumes of material up to the order of one-half litre per day (depending upon the sedimentation constant of the material), is not convenient for centrifuging greater quantities. Larger quantities can be centrifuged by means of the tubular vacuum-type ultracentrifuge (Beams, 1930 a–40). In this machine the material to be centrifuged enters the top of a vertical spinning tube through a hollow shaft at a continuous rate. The separation takes place as the material flows from top to bottom of the vertical spinning tubular rotor and emerges in a light and a heavy fraction through two coaxial tubes which serve as the lower shaft. The tubular rotor contains a solid core to confine, to the strongest part of the field, the material to be centrifuged. Since the material enters and leaves the rotor along the axis and the rotor is spun in a vacuum, the power required to spin it is very small. With this machine the amount of a given material that can be concentrated per unit of time is about one hundred times greater than in the case of the quantity-type centrifuge.

The air-driven vacuum-type centrifuge has also been applied to the concentration of isotopes (Beams, 1937, 38 b, 40; Beams and Skarstrom, 1939; Humphreys, 1939). In general, two methods have been used in this work. The first is the evaporative centrifuge method suggested by Mulliken (1922). In this method the material in the liquid state is contained in a layer around the cylindrical inside wall of a hollow vacuum-type rotor. This material slowly evaporates and diffuses in the vapour state to the axis against the centrifugal field. During this diffusion the separation takes place. The material flows out of the rotor through a hollow shaft and is collected in a series of fractions. The first fractions show a concentration of the lighter isotopes, while the last fractions show a concentration of the heavy isotopes. The experiments show that the theory (equation (12)) holds unless, of course, the material is removed from the rotor too rapidly for approximate equilibrium to be established. The second centrifuge method of concentrating isotopes employs the tubular vacuum-type centrifuge (Beams, 1937, 38 b, 40). The material in the gaseous or vapour state enters at the top of the vertical spinning tubular centrifuge through a hollow shaft. As it flows down along the length of the tube, the separation occurs, with the heavier fraction moving toward the periphery and the lighter fraction toward the axis. Each of these fractions flows out of the machine through one of two channels formed by two coaxial tubes comprising the lower shaft. This method also gives the approximate separation expected by the theory when the material is not forced through the machine too rapidly. In addition to the concentration of isotopes, this type of machine has been used for the separation of N_2 and CO_2, N_2 and O_2, etc. (Beams, 1937, 38 b, 40).

§7. ELECTRICALLY DRIVEN MAGNETICALLY SUPPORTED VACUUM-TYPE ULTRACENTRIFUGE

The electrically driven magnetically supported vacuum-type ultracentrifuge was developed primarily to give accurate automatic speed control over long periods of time and to avoid the necessity for a supply of compressed air (Beams and Snoddy, 1937 ; Beams, 1937, 38b, 40 ; Skarstrom and Beams, 1940). This centrifuge is self-balancing and utilizes the same mechanical principles as the vacuum-type air-driven air-supported centrifuge. The larger rotor spins in a vacuum and hence has the same temperature control as the vacuum-type centrifuges previously described. Only the drive and support are different. The rotating members are driven by an induction motor, the field of which is actuated by a high-frequency power supply. The rotating members are supported (except for a very minute part of the weight which is taken by a small thrust bearing) by the upward pull of an electromagnet on an iron or steel rod fastened to the shaft and coaxial with it. This type of magnetic support is almost friction-free if constructed properly, because the magnetic field is symmetrical across the end of the rod and no appreciable electromagnetic drag exists. The armature of the induction motor is a cylindrical rod of strong steel and is water-cooled, especially during the starting period, when the " slip " is large. The motor and support are situated either above or below the vacuum chamber, i.e., they are situated in the same position as the air turbine in the machines described above. The rotational speed is held constant ($\pm 0{\cdot}5$ r.p.s. for days at a speed of 1100 r.p.s.) by a speed control operated by the " slip ", i.e., the difference between the driving frequency and the centrifuge speed. The electrical motor is more efficient than the air drive, has a good starting torque, is almost entirely automatic, and operates from the 110-volt a.c. lines (Skarstrom and Beams, 1940). Since the armature of the motor is small in diameter in comparison with the large rotor which it spins inside the vacuum chamber, the only factor which limits the rotor speed is the strength of the rotor material. These electrical motors have been made to operate up to about 3000 r.p.s., but this is not by any means the upper limit.

Recently an effective method of spinning rotors to high speeds ***in vacuo***, which utilizes the magnetic suspension of F. T. Holmes, has been developed. (Holmes, 1937 ; Holmes and Beams, 1937 ; Beams, 1937, 38 b, 40 ; MacHattie, 1941 ; Smith, 1941). The rotating members are attached to a vertical steel rod or ball which is suspended in the coaxial field of one or more solenoids. When two solenoids are used, the magnitude of the steady d.c. current passed through the larger solenoid is not quite sufficient to lift the rotating members. The remaining weight is supported by the second solenoid, the current through which is controlled by the amount of light falling upon a photoelectric cell. The intensity of the incident light striking the photoelectric cell is increased as the centrifuge rotor moves upward and decreased as it moves downward, in such a way that the rotor is maintained at a predetermined height. This stabilizes the

vertical motion while the geometry of the field stabilizes the horizontal motion of the rotor. In some arrangements the current in the supporting solenoid (and hence the height of the rotor) is regulated by making the rotor one plate of a condenser (Smith, 1941). The change in capacity of this condenser changes the current in the solenoid. In another arrangement by MacHattie (1941) the change in characteristics of a small coil carrying a radio-frequency current, as the rotor approaches or recedes, is made to regulate the current in the supporting solenoid. The supporting magnetic field is symmetrical across the spinning rod so that there is no appreciable electromagnetic drag or friction opposing axial rotation. Also, the rotating members are not mechanically connected to anything and can be completely surrounded by a glass container. This allows the rotor to be degassed by heating and makes it possible to obtain a very high vacuum. The rotors have been spun by rotating electrostatic and by rotating magnetic fields. A small rod which was spun to 1000 r.p.s. in a good vacuum, and the driving torque removed, required about 1000 sec. to lose one r.p.s. A steel ball 3/32" diameter was spun to $6{\cdot}5 \times 10^6$ r.p.m., giving a centrifugal field of about 57×10^8 g (MacHattie, 1941).

In practically all of the types of centrifuges described above, the mechanical strength of the rotor limits the rotational speed attainable. For a smooth solid rotor of a given shape (for example, one of uniform strength) the bursting speed depends upon the yield-point divided by the density of the material. If the rotor carries a load or has been bored or pierced with channels, as is required in many experiments, local stresses may rise to very high values over comparatively small regions. If the material is plastic (has a large elongation), when the local stresses exceed the yield-point of the material, it flows and distributes the local stresses over a larger volume (as well as increases the strength of the material locally) and the rotor does not explode. On the other hand, if the material has a small elongation, the local stresses start cracks which cause the rotor to explode. For this reason some investigators have had disastrous results with rotors made of very strong hard steel (Pollock and Collie, 1937 ; Beams, 1937, 38 b, 40 ; Svedberg and Pedersen, 1940). For many purposes, the alloy steels such as SAE 4130 X, heat-treated to a Brinell hardness between 400 and 450 (yield-point 200,000 lb./in^2), or the tough nickel alloy steels, etc., are preferable to the stronger but harder steels. Some of the aluminium and magnesium alloys have been used extensively for centrifuge rotors. They have a yield-point/density ratio almost as large as the alloy steels mentioned above. Duralumin ST 14 has had very wide use because of its favourable elastic properties and its ready machinability.

§8. APPLICATIONS

High-speed rotors have been applied to many different fields of research. As a matter of fact they should be classed with that rapidly growing group of physical apparatus which has made possible so much recent borderline research

between the sciences. Already the applications of high-speed centrifuges are too numerous to permit a reasonably complete review in this *Report*. Instead, a few results will be mentioned to illustrate the types of problems that have been undertaken and briefly to call attention to some data usually considered outside the province of physics, yet which may be of interest to physicists.

From the standpoint of biology and medicine, one of the most important applications of the ultracentrifuge is in the purification in quantities of the naturally occurring substances such as viruses, hormones, etc., without deactivating them (Elford, 1938; Severinghaus, Levin and Chiles, 1938; Wyckoff, 1937 a–38 b; Bauer and Pickels, 1936–40). These substances, which often occur in very dilute solution, are known to influence critically the well-being of plants and animals. In fact, their existence was usually determined by their biological effects. Consequently their purification and concentration by the high-speed centrifuge has put into the hands of the experimenter concentrated specimens whose properties can be investigated. For example, Bauer and Pickels (1936–40) obtained a ten-thousand-fold change of concentration by one centrifuging of yellow-fever virus, while Wyckoff (1937 a–38 b) and his associates have concentrated several of the larger viruses until they crystallized by one centrifuging. Incidentally, it might be of interest to note that crystallization of some of these viruses any number of times apparently does not affect their biological activity.

In their earlier work Svedberg (1928–39) and his collaborators, contrary to the belief at that time, discovered that certain proteins such as haemoglobin, egg albumen, etc., consisted of homogeneous molecular species with high molecular weights rather than colloids (particles of various masses made up of ordinary small molecules). Subsequently a large number of substances composed of these giant molecules has been studied in the ultracentrifuge, by many different workers, and their molecular weights, sizes and shapes determined. Not all of these naturally occurring substances, such as keratin, fibroin, and myosin, are composed of homogeneous molecular species, but at least most of the easily soluble proteins have definite molecular weights. Consequently the high-speed centrifuge has made available pure substances with molecular weights which may exceed one hundred million (Wyckoff, 1937 a–38 b). Clearly such substances should possess novel physical properties, especially in cases where the molecules are elongated. Several physical properties of a few of the viruses, such as streaming double refraction, viscosity, Kerr electro-optical effect, spontaneous double refraction, etc., have been studied, and most interesting results obtained (Lauffer, 1939; Lauffer and Stanley, 1939; Svedberg and Pederson, 1940). Also some of these giant molecules show somewhat surprising chemical properties (Wyckoff, 1937 a–38 b; Brohult, 1940; Svedberg and Pedersen, 1940). The analytical centrifuge shows that many of the protein molecules are dissociated or associated, usually reversibly, by such weak agents as small changes in *p*H, changes in salt concentration, or even changes in concentration of the protein itself. In some cases, such as the respiratory proteins, which have been extensively studied, giant molecules dissociate

into sub-multiples. For example, when a solution of haemoglobin is made very dilute, the molecules split into halves. In the case of haemocyanin (*Helix Pomatia*) solution near the isoelectric point, whole molecules (molecular weight = $8{\cdot}9 \times 10^6$) exist alone provided the salt concentration is not too high. When the solution is made weakly acid, half-molecules appear. If it is made alkaline, half-molecules first appear, followed by molecules of one-eighth the original molecular weight, etc. Within the region of *p*H 4 to *p*H 9 the dissociation is reversible, and the whole molecules will be re-formed at the isoelectric point. Outside the *p*H region (4–9) non-homogeneous components appear and will not completely re-associate. It is interesting to note that in this case the whole, the half, and the one-eighth molecules all have approximately the same length (Brohult, 1940). Svedberg (1928–39) has pointed out that the molecular weights of the soluble proteins which are composed of homogeneous molecular species are integral multiples of a unit weight (17,600) ; he has, in fact, divided most of these proteins into eleven classes with molecular weights equal to integers containing some power of 2 or 3 multiplied by 17,600. The rule does not fit exactly in all cases, but undoubtedly indicates a regularity in the masses of the very large virus proteins, which he thinks cannot be due to chance.

The fact that many substances have molecular weights well over 5×10^7 raises interesting theoretical questions as to why they are stable or why they are all exactly the same size. Usually they are assumed to have the same mass because each molecular species sediments in the ultracentrifuge with a sharp sedimentation boundary. However, at present too much confidence should not be placed in this conclusion because possible variations in mass of 10^6 which are not observable may occur in some instances. There is need for more precise observations on the homogeneity of some of these giant or macro-molecules.

McFarlane (1935, 38), Pedersen (Svedberg and Pedersen, 1940) and others have used the ultracentrifuge to study differences between normal and pathological sera and have found in many cases that the changes are pronounced. In some diseases new sedimentation boundaries appeared along with changes in the amounts of the various components such as the globulin or albumen.

In addition to the proteins mentioned above, the molecular weights and sizes of many other substances have been studied by means of the ultracentrifuge. Lamm (1929 a–37) and Beckmann and Landis (1939) have determined the masses of the starch molecules under various conditions. Nichols, Kraemer, and others * have studied a large number of materials used industrially, such as neoprene, rubber and cellulose. In some cases they have been able to correlate the sizes and shapes of the molecules of the substance with its physical properties. Many of the synthetic compounds show no sharp sedimenting boundaries, indicating that they are polydisperse, i.e., they are made up of particles whose weights vary over considerable ranges. Fortunately, the ultracentrifuge data give not only the masses of the particles but their number or concentration as well. In some solutions the rate of sedimentation in the centrifuge is influenced rather

* See Svedberg (1928–39) and Svedberg and Pederson (1940).

critically by the concentration. This is true especially in the case of the polystyrenes dissolved in organic solvents, where the rate of change of the sedimentation constant increases as the dilution is increased. Such behaviour is thought to be due to the entanglement, or interference with each other, of the long thread-like molecules. Many natural substances, as well as those made synthetically, which have long thread-like molecules, are polydisperse. It should be noted that in addition to giving the particle-weight distribution, the ultracentrifuge can distinguish between mixtures of molecules and chemical equilibrium (Pedersen, 1934; Svedberg, 1928–39).

When electrolytes are placed in a gravitational field, electrical potentials are developed between two electrodes at different heights (MacInnes, 1939). The phenomenon is magnified in the centrifuge, as has been shown by Tolman (1911), MacInnes, and others. According to theory, the potential difference between centre and periphery of the centrifuge should be a maximum when the centrifuge first reaches full speed and then decrease with time until equilibrium is established in the centrifuge cell. From these measurements it is possible to determine the transference numbers. By comparing the molecular weight of an electrolyte obtained by the equilibrium method of centrifuging with that derived from chemical data, Pedersen (1934), Drucker (1937) and others have been able to obtain the activity coefficients.

Interesting investigations have been carried out with the high-speed centrifuge on the sedimentation occurring inside a living cell (Harvey, 1932–38; Beams (H.W.) and King, 1934–38). Estimates have been made of the relative density of the various components of the cell, the " viscosity," and " surface tension " of the protoplasm and the elasticity of the cell membranes. In some cases certain components of the cell were thrown through the cell wall, yet the cell continued to function. In other cases, after milder centrifuging, the cell developed abnormally or died. H. W. Beams and King (1934–38) subjected fertilized eggs of *Ascaris suum* to fields of 4×10^5 gravity for an hour. The eggs were then removed from the centrifuge and observed under a microscope. At first the contents of the cell were stratified into definite layers. However, twelve hours later they had lost their stratification and many of them continued to live.

As mentioned earlier, the high-speed centrifuge has been used to concentrate the isotopes of several elements. Most of this centrifuging was done with the substances in the vapour or gaseous state. The separation of gaseous mixtures such as O_2 and N_2 or N_2 and CO_2 has also been studied. When the rate of flow through the centrifuge was not too great, the simple theory (equation (12)) was found to hold. In addition to the centrifuging method, the separation of gaseous mixtures has been accomplished by the use of a high-speed rotor of special design used as a velocity selector (Beams, 1937, 38 b, 40).

Fortunately, most of the high-speed rotors discussed in this article are stable enough to carry a polished mirror. Since the bursting strength of the rotor is the only factor which limits rotational speed, the construction of very high-speed rotating mirrors is possible. Such high-speed mirrors have been mostly

used to study phenomena which occur in very short intervals of time (Chiles, 1937 Beams, 1930 a–40).

GENERAL REFERENCE

SVEDBERG and PEDERSEN, 1940. *The Ultracentrifuge* (Oxford). (See for other references.)

OTHER REFERENCES

ANDERSSON, 1939. *Nature, Lond.*, **143**, 720.
ARCHIBALD, 1938 a. *Phys. Rev.* **53**, 746.
ARCHIBALD, 1938 b. *Phys. Rev.* **54**, 371.
BAUER and PICKELS, 1936. *J. Exp. Med.* **64**, 503.
BAUER and PICKELS, 1937. *J. Exp. Med.* **65**, 565.
BAUER and PICKELS, 1940. *J. Exp. Med.* **71**, 703.
BAWDEN and PIRIE, 1938. *Nature, Lond.*, **141**, 513.
BEAMS, H. W. and KING, 1934. *Anat. Rec.* **59**, 363, 395.
BEAMS, H. W. and KING, 1935 a. *Proc. Roy. Soc.* B, **118**, 264.
BEAMS, H. W. and KING, 1935 b. *Nature, Lond.*, **135**, 232.
BEAMS, H. W. and KING, 1936 a. *Biol. Bull.* **71**, 188.
BEAMS, H. W. and KING, 1936 b. *Science*, **84**, 138.
BEAMS, H. W. and KING, 1937. *Biol. Bull.* **73**, 99.
BEAMS, H. W. and KING, 1938. *Biol. Bull.* **75**, 189. (See Beams, H. W. and King, 1934–38 for other references.)
BEAMS, J. W., 1930 a. *Rev. Sci. Instrum.* **1**, 667.
BEAMS, J. W., 1930 b. *Phys. Rev.* **35**, 24 ; **36**, 997.
BEAMS, J. W., 1933. *Phys. Rev.* **44**, 803.
BEAMS, J. W., 1935. *J. Opt. Soc. Amer.* **25**, 48.
BEAMS, J. W., 1937. *J. Appl. Phys.* **8**, 795.
BEAMS, J. W., 1938 a. *Rev. Sci. Instrum.* **9**, 413.
BEAMS, J. W., 1938 b. *Rev. Mod. Phys.* **10**, 245.
BEAMS, J. W., 1939. *Science*, **89**, 543.
BEAMS, J. W., 1940. *Science in Progress*, **2**, chapt. 9, 232 (Yale University Press). (See Beams, J. W., 1930 a–40, for other references.)
BEAMS, J. W. and PICKELS, 1935. *Rev. Sci. Instrum.* **6**, 299.
BEAMS, J. W. and SKARSTROM, 1939. *Phys. Rev.* **56**, 266.
BEAMS, J. W. and SNODDY, 1937. *Science*, **85**, 185.
BEAMS, J. W., LINKE and SOMMER, 1938. *Rev. Sci. Instrum.* **9**, 248.
BEAMS, J. W., WEED and PICKELS, 1933. *Science*, **78**, 338.
BECKMANN and LANDIS, 1939. *J. Amer. Chem. Soc.* **61**, 1495 and 1504.
BJÖRNSTAHL, 1939. *Rev. Sci. Instrum.* **10**, 258.
BROHULT, 1940. *Nova Acta Reg. Soc. Sci. Upsal.* (4), **12**, no. 4.
CHAPMAN, 1919. *Phil. Mag.* **38**, 182.
CHILES, 1937. *J. Appl. Phys.* **8**, 622.
DAVIS, 1936. *Rev. Sci. Instrum.* **7**, 96.
DRUCKER, 1937. *Z. Phys. Chem.* **180**, 359 and 378.
ELFORD, 1938. *Handbuch der Virusforschung* (Springer). (See for other references.)
FAXÉN, 1929. *Ark. Mat. Astr. Fys.* **21** B, no. 3.
FAXÉN, 1936. *Ark. Mat. Astr. Fys.* **25** B, no. 13.
GARMAN, 1933. *Rev. Sci. Instrum.* **4**, 450.
GIRARD and CHUKRI, 1933. *C.R. Acad. Sci., Paris*, **196**, 327.
HARVEY, 1932. *J. Franklin Inst.* **214**, 1.
HARVEY, 1934. *Biol. Bull.* **66**, 48.
HARVEY, 1938. *J. Appl. Phys.* **9**, 68.
HENRIOT and HUGUENARD, 1925. *C.R. Acad. Sci., Paris*, **180**, 1389.
HENRIOT and HUGUENARD, 1927. *J. Phys. Radium*, **8**, 443.
HOLMES, 1937. *Rev. Sci. Instrum.* **8**, 444.
HOLMES and BEAMS, J. W., 1932. *Nature, Lond.*, **140**, 30.
HUMPHREYS, 1939. *Phys. Rev.* **56**, 684.
JOLY, 1921. *Phil. Mag.* **41**, 818.

JOLY and POOL, 1920. *Phil. Mag.* **39**, 372.
KAHLER, 1938. *Rev. Sci. Instrum.* **9**, 257.
KRAEMER, 1941. *J. Franklin Inst.* **231**, 1.
KRAEMER and LANSING, 1934. *Nature, Lond.*, **133**, 870.
KRAEMER and LANSING, 1935 a. *J. Phys. Chem.* **39**, 153.
KRAEMER and LANSING, 1935 b. *J. Amer. Chem. Soc.* **57**, 1369.
KRAEMER and LANSING, 1936. *J. Amer. Chem. Soc.* **58**, 1471.
LAMM, 1929 a. *Ark. Mat. Astr. Fys.* **21** B, no. 2.
LAMM, 1929 b. *Z. phys. Chem.* **143** A, 177.
LAMM, 1937. *Nova Acta Reg. Soc. Sci. Upsal.* (4), **10**, no. 6.
LAUFFER, 1939. *J. Amer. Chem. Soc.* **61**, 2412.
LAUFFER and STANLEY, 1939. *Chem. Rev.* **24**, 303.
LINDEMANN and ASTON, 1919. *Phil. Mag.* **37**, 523.
MCBAIN, 1939. *Chem. Rev.* **24**, 289.
MCBAIN and LEYDA, 1938 a. *Nature, Lond.*, **141**, 913.
MCBAIN and LEYDA, 1938 b. *J. Amer. Chem. Soc.* **60**, 2998.
MCFARLANE, 1935. *Biochem. J.* **29**, 407, 660 and 1175.
MCFARLANE, 1938. *Nature, Lond.*, **142**, 1023.
MACHATTIE, 1941. *Phys. Rev.* **59**, 468.
MACINNES, 1939. *Principles of Electrochemistry* (Reinhold).
MCINTOSH and SELBIE, 1937. *Brit. J. Exp. Path.* **18**, 162.
MASKET, 1940. *Phys. Rev.* **57**, 1056.
MASON and WEAVER, 1924. *Phys. Rev.* **23**, 412.
MULLIKEN, 1922. *J. Amer. Chem. Soc.* **44**, 1033 and 1729.
NICHOLS, 1928. *Colloid Symp. Monograph*, **6**, 287.
NICHOLS, 1931. *Physics*, **1**, 254.
PEDERSEN, 1934. *Z. phys. Chem.* **170** A, 41.
PHILPOT, 1938. *Nature, Lond.*, **141**, 283.
PICKELS, 1936. *Science*, **83**, 471.
PICKELS, 1938. *Rev. Sci. Instrum.* **9**, 354.
PICKELS and BEAMS, J. W., 1935. *Science*, **81**, 342.
POLLOCK and COLLIE, 1937. *The Engineer*, **166**, 102.
SEVERINGHAUS, LEVIN and CHILES, 1938. *Endocrinology*, **23**, 285.
SHAPIRO and BUTT, 1937. *Rev. Sci. Instrum.* **8**, 35.
SKARSTROM and BEAMS, J. W., 1940. *Rev. Sci. Instrum.* **11**, 398.
SMITH, 1941. *Rev. Sci. Instrum.* **12**, 15.
SNODDY and BEAMS, J. W., 1937. *Science*, **85**, 273.
SOYTEN, 1936. *Proc. Phys.-Math. Soc. Japan*, (3) **18**, 9.
SOYTEN, 1937. *Proc. Phys.-Math. Soc. Japan*, (3) **19**, 1094.
STAUFFACHER, BEAMS, J. W. and SNODDY, 1941. *Phys. Rev.* **59**, 468.
SVEDBERG, 1928. *Colloid Chemistry* (Chem. Cat. Co.).
SVEDBERG, 1937. *Nature, Lond.*, **139**, 1051.
SVEDBERG, 1938. *Ind. Engng Chem.* (Anal. Edn.), **10**, 113.
SVEDBERG, 1939. *Proc. Roy. Soc.* B, **127**, 1. (See Svedberg, 1928–39, for other references.)
TISELIUS, PEDERSEN and SVEDBERG, 1937. *Nature, Lond.*, **140**, 848.
TOLMAN, 1911. *J. Amer. Chem. Soc.* **33**, 121.
UREY, 1940. *Rep. Progr. Phys.* **6**, 48.
WEAVER, 1926. *Phys. Rev.* **27**, 499.
WILSON, 1940. *Phys. Rev.* **58**, 209.
WYCKOFF, 1937 a. *Proc. Amer. Phil. Soc.* **77**, 445.
WYCKOFF, 1937 b. *Science*, **86**, 92 and 311.
WYCKOFF, 1937 c. *J. Biol. Chem.* **121**, 219 ; **122**, 239.
WYCKOFF, 1938 a. *J. Biol. Chem.* **124**, 585.
WYCKOFF, 1938 b. *Cold Spring Harbor Symposia*, **6**. (See Wyckoff, 1937–38 b, for other references.)

APPLICATIONS OF SPECTROSCOPY TO COMBUSTION

By A. G. GAYDON,*
Imperial College, London

§ 1. INTRODUCTION

As this subject is being dealt with for the first time in these *Reports*, it seems desirable to touch briefly on the historical side and to cover the groundwork fairly thoroughly, indicating the present trend of progress and lines of future development of the applications of spectroscopy to combustion without dealing in detail with all the latest original papers.

It is convenient to divide the subject into four main sections. The first deals with the more obvious approach to the subject, the visible and ultra-violet *emission spectra* of flames, explosions, cold flames, etc. The second section is devoted to *absorption spectra*, and the third section to the *infra-red* region. The last section, on the subject of *energy considerations*, is concerned with the large number of indirect but extremely fruitful applications of spectroscopy to combustion problems, such as the calculation of dissociation energies, specific heats, quantitative considerations of the lifetimes of activated molecules, and temperature measurement.

§ 2. EMISSION SPECTRA

One of the first spectra studied was that of the candle flame, and as long ago as 1802 Wollaston described the banded spectrum which was later, in 1857, mapped by Swan, by whose name it is now usually known. Burch (1886), Smithells (1901) and many others observed the spectra emitted by various flames. Such observations were, however, of little help to the theory of combustion until theoretical developments in the interpretation of the rotational structure of diatomic band spectra and the systematization of such spectra by Hund (1926), Mulliken (1930–32) and others enabled the emitting molecules to be identified with some certainty. Such developments and systematization have been described by Jevons (1932).

The spectra observed in flames and combustion processes are in many cases emitted by unstable intermediary compounds and radicals such as OH, CH, C_2, CN and NH, and the discovery of the existence of such intermediaries has, since the work of Bonhoeffer and Haber (1928), had a marked effect on combustion theory and chemical kinetics generally. In using the empirical knowledge

* Leverhulme Research Fellow.

that such radicals are present in flames, it must, however, be remembered that a very small proportion of excited molecules will give a strong emission spectrum, and that these radicals are frequently present in such very small numbers that they cannot readily be observed by their absorption spectrum. Also, very many molecules which may be present in the flame may give no evidence of their presence spectroscopically because either they are not electronically excited, or their electronic levels are so spaced that no emission spectrum is obtainable in the region of the spectrum being studied; thus we know that molecules of CO, O_2 and N_2 must be present in flames in large numbers, but such high excitation energies are required to excite their spectra that the relatively mild conditions existing in flames are insufficient to produce this excitation. It is a general property of the spectra excited in combustion processes that only electronic levels of low energy are involved, and the spectra almost invariably correspond to transitions to the lowest electronic state. Flame spectra also differ from spectra excited in other sources, such as discharge tubes and arcs, in that the excitation in flames appears to be due largely to collisions between particles of atomic or molecular dimensions or to chemiluminescence, while in discharge tubes and arcs the excitation is mostly by electron impact. Thus some flame spectra, such as that of the carbon-monoxide flame and the so-called ethylene flame bands, have not been observed in other sources, while others, like the spectra of CaI, SrBr (Hedfeld, 1931) and NiH (Gaydon and Pearse, 1935), are only observed with difficulty in other sources, although appearing readily in flames.

The hydrogen flame

The spectrum of burning hydrogen consists almost entirely of a system of bands, degraded to the red, with open rotational structure. The strongest band has an outstanding head at 3063·6 A., and there are heads of weaker sequences of bands at 3428, 2811 and 2608 A. This band system is now known to be due to the radical OH, following the rotational analysis of the bands and calculation of the moment of inertia of the molecule by Watson (1924) and Jack (1927). The knowledge that such radicals were present in the flame led Bonhoeffer and Haber (1928) to postulate, for the combustion, a chain mechanism of the type

$$H + O_2 + H_2 = H_2O + OH + 97{\cdot}9 \text{ kcal.}$$
$$OH + H_2 = H_2O + H + 16{\cdot}3 \text{ kcal.}$$

Several other chain mechanisms have been proposed for the reaction between oxygen and hydrogen, some of which involve the occurrence of the unstable intermediary HO_2, but the production of OH at certain stages is generally accepted. Kondratjew and Ziskin (1937 b) have made quantitative measurements of the relative intensities of the (0, 0), (1, 0) and (2, 1) bands of OH and find that the intensities of these bands are independent of the flame temperature and are proportional to the transition probabilities, the populations in the levels $v' = 0$, 1

and 2 not being governed by the usual Maxwell-Boltzmann distribution. From this they conclude that the excitation of the OH radicals is not thermal, but arises from the chemical reaction

$$H + H_2 + O_2 = H_2O + OH + 102 \text{ kcal.}$$

Allowance being made for the temperature of the flame (1000° K.), the available energy is then 110·5 kcal. Excitation to the $v' = 0$, 1, 2 and 3 levels of electronically excited OH requires energies of 92·1, 100·5, 108·5 and 116·0 kcal. respectively. Thus the OH radicals may be formed in an electronically excited state by the reaction process given above; the sudden falling off in the intensity of the (3, 2) band which is observed in the flame is accounted for by the reaction energy being insufficient to excite the $v' = 3$ level. The excited OH radicals may (Farkas, 1931; Herzberg, 1939) result in chain branching by the mechanism

$$OH^* + H_2 = OH + H + H - 10{\cdot}4 \text{ kcal.}$$

$$\text{or} \quad OH^* + H_2 = O + H + H_2 - 6{\cdot}6 \text{ kcal.}$$

if the temperature is sufficient to provide the small activation energies required for these processes. Other quantitative measurements on the OH radiation from flames have been made by Kondratjew and Ziskin (1937 a, 37 b), from which they conclude that only one excited OH radical is formed for every million H_2O molecules produced.

The hydrogen flame shows, in addition to the strong OH bands, a very weak continuous background. Kondratjew and Ziskin have suggested that this may be connected with the oxygen afterglow which appears in the afterglow of water (Lavin and Stewart, 1929) when oxygen is present. Kitagawa (1936) has observed a complex system of bands in the orange and red from a flame of oxygen burning in hydrogen, and has tentatively identified these with the vibration-rotation spectrum of water. The confirmation of the presence of such highly activated water molecules would be of considerable interest to the combustion theory.

Hydrocarbon flames

The inner cone of the ordinary Bunsen-type flame of burning hydrocarbons, such as methane, ethylene, acetylene and higher hydrocarbons and of burning ethers, gives a rather striking banded spectrum. In the visible region the most prominent feature is the system of Swan bands, with sequence heads at 6191, 5635, 5165 (0, 0), 4737 and 4382 A. These bands, which are degraded to the violet, form a strikingly simple system. Their emitter was for long in doubt, but the study of the rotational structure, revealing alternation of intensities, which is known to be characteristic of homonuclear molecules, and experimental work by Pretty (1928), finally established that the emitter was C_2.

In the violet at about 4315 and 3900 A. there are strong bands with open rotational structure. The stronger band at 4315 has a strong head at 4312·5 A. and a weaker head at 4315 A. and is degraded to the violet, while the 3900 band

shows heads at 3889 and 3872 A. and is degraded to the red. These bands are due to the radical CH, and their structure has been studied by Kratzer (1924), Mulliken (1927) and Grenat (1931). A few weaker bands of these systems are occasionally observed with heads at 4890, 4025 and 3628 A. Relatively hot flames also show a band due to CH (Hori, 1930) at 3143 A., which is degraded to shorter wave-lengths. The OH bands also appear strongly in hydrocarbon flames; and the outer cone shows, in addition to OH, some of the continuous and faintly banded structure due to burning carbon monoxide.

In addition to the above strong band systems, the inner cones of hydrocarbon flames show a weak system of bands in the near ultra-violet. This system has hitherto been known as the " ethylene flame bands ", following the detailed study of the structure of the bands in the ethylene flame by Vaidya (1934), but they occur in the flames of all hydrocarbons, including methane and acetylene, and may preferably be called " the hydrocarbon flame bands ". The system consists of a large number of bands, degraded to the red, among which the heads at 3589, 3503, 3377, 3299 and 3015 A. are relatively strong. The bands have been variously attributed to HCO (Vaidya, 1934, 41), CH_2 (Bell, 1937) and CHOH, an isomer of formaldehyde (Smith, 1940), the first-named being perhaps the most likely.

The detailed knowledge of the spectra of burning hydrocarbons, which has been obtained, has, however, been of surprisingly little help to the theory of combustion. Recent observations, of the hydrocarbon flame bands in the outer cones of chilled hydrocarbon flames by Gaydon (1942), combined with chemical tests on the interconal gases, seem to suggest that the hydrocarbon flame bands may be associated with the decomposition of peroxidic susbtances; if this is substantiated by further work now in progress it may contribute materially to the controversy existing between the rival peroxidation and hydroxylation theories of the mechanism of hydrocarbon combustion.

Cool flames

Most of the higher hydrocarbons, ethers and aldehydes are capable of slow combustion at relatively low temperatures, and this slow combustion is frequently accompanied by either a cool flame or luminescence of the reacting gases. These cool flames are always of low luminosity, and very long exposures on high-aperture instruments are necessary to record their spectra. Emeléus (1926, 29 a) has examined the spectra of the cool flames of diethyl ether, hexane, acetaldehyde and propaldehyde and finds that they all show the same banded spectrum in the near ultra-violet. This spectrum has been identified by Pearse (see Ubbelohde (1935 a)) and Kondratjew (1936) with the fluorescence spectrum of formaldehyde. Kondratjew points out that the low temperature of such flames excludes the possibility of the thermal excitation of the formaldehyde molecules; it is possible that these molecules act as recipients of energy liberated in some branch of the reaction chain, or alternatively, that they are formed in an

electronically excited state. This discovery of the formation of electronically excited formaldehyde in slow combustion has been discussed by Ubbelohde (1935 a) in a paper on the peroxidation theory of hydrocarbon combustion; von Elbe and Lewis (1937), who favour a step-wise breaking down of the higher hydrocarbons through peracids and lower aldehydes, postulate

$$CH_3CO + O_2 = \underset{\text{excited formaldehyde}}{HCHO^*} + CO + OH$$

as one of the last reactions in all these combustion processes, there being an energy of 110 kcal. available from this reaction to excite the newly-formed formaldehyde molecules, which then emit this energy as radiation, giving the cool flame.

The cool flame of carbon disulphide has been examined by Emeléus (1926) and by Fowler and Vaidya (1931), who found that its spectrum differed from that of the normal flame only in the inclusion of a few weak bands of CS, the strongest features being the bands of S_2 and SO. Other cool flames or phosphorescent glows studied by Emeléus and his colleagues include those of the slow oxidation of phosphorus (Emeléus and Downey, 1924; Emeléus and Purcell, 1927), phosphorus trioxide (Emeléus, 1925) and arsenic (Emeléus, 1927, 29 b), and of the reactions of selenium oxide with alcohols and with ammonia (Emeléus and Riley, 1933).

Other low-temperature flames which have been investigated spectroscopically are the so-called "atomic flames". These are produced when atomic oxygen or atomic hydrogen, obtained from a discharge-tube, is mixed with an inflammable vapour. Preliminary investigations of the spectra of these atomic flames were made by Harteck and colleagues, and recently Geib and Vaidya (1941) have studied the flames of hydrocarbons, including ethylene and acetylene, burning in atomic oxygen. They found that the spectra of these atomic flames were essentially similar to those of the inner cones of the corresponding ordinary hot flames. This would appear to indicate that the combustion processes in the atomic flames differ from those in the cool flames, but resemble more those in hot flames. It might at first be thought that the similarity of the spectra of the atomic and hot flames indicated that atomic oxygen was of importance in the combustion of hydrocarbons (Norrish's modified hydroxylation theory), but this is disproved by the fact that methane and atomic oxygen will not produce a flame ; nor will hydrogen or carbon monoxide mixed with atomic oxygen.

Although the spectra of the atomic flames are generally similar to those of the hot flames, there are certain differences in detail. Thus the hydrocarbon flame bands were observed by Geib and Vaidya to occur strongly in the atomic flame of acetylene, but only weakly in the flame of ethylene, although the reverse is true of the hot flames. These hydrocarbon flame bands were not found in the flame of benzene reacting with atomic hydrogen, and this supports the theory

that the emitter of these bands is a molecule such as HCO which contains oxygen. The formaldehyde—atomic-oxygen flame showed only OH bands, while the methyl-alcohol—atomic-oxygen flame showed the same spectrum as the cool flame of ether, with strong OH and weak CH bands superposed.

The carbon-monoxide flame

The spectrum of the flame of carbon monoxide burning in air or oxygen appears at first glance to be a continuum stretching from the visible region to far in the ultra-violet, with a maximum of intensity in the violet. Closer inspection, however, shows that this spectrum is, as was pointed out by Burch in 1886, of a banded character, the bands showing up more clearly when the gas is burnt at reduced pressure. The system is very complex, consisting of a very large number of narrow bands, often occurring in close pairs, and usually without sharp heads. With large dispersion the system shows indications of line structure. Unless very great care is taken to dry the gases, the bands of OH at 3064 A. always appear strongly in the ordinary flame, although when the gas is burnt at high pressure the OH bands appear less readily. This fact was observed by Weston (1925 a, 25 b), whose experimental work also established that the spectrum was produced from pure dry carbon monoxide burning in oxygen, and a systematization of the band spectrum, made by Fowler, is included in Weston's paper. Kondratjew (1930) established the occurrence of intervals of around 600 $cm.^{-1}$ in the band spectrum, and provisionally identified this interval with the transverse vibrational frequency of CO_2, which is 667 $cm.^{-1}$.

Gaydon (1940) discusses the possible origin of the carbon monoxide flame spectrum and notes that the choice of emitter can be limited to a polyatomic oxide of carbon, of which carbon dioxide is the most likely. The absorption spectrum of the suboxide C_3O_2 shows some resemblance to the flame bands, but this molecule is improbable, as the emitter, on other grounds. A peroxide CO_3 is also a possibility, but no evidence for the presence of this has been obtained from experiments on the slow combustion of carbon monoxide. Carbon dioxide in gaseous or liquid form is transparent throughout the visible and quartz ultra-violet regions, and the flame bands are not obtained from CO_2 in discharge tubes. Comparison with the Schumann-Runge bands of O_2 suggests, however, that, if there is a big difference in the shape or size of the CO_2 molecules in the two electronic states, then the flame bands may possibly form part of the absorption band system of CO_2, which is known to exist below 1700 A. A consideration of the electronic energy levels of CO_2 shows that normal CO_2 is not formed from normal CO and O_2 (Schmid, 1936), and Gaydon therefore suggested that an electronic re-arrangement of the newly-formed CO_2 molecules must occur after the initial combustion process, and that, if this re-arrangement results in a considerable change in the shape or size of the molecule, the molecules formed after the combustion will be "vibrationally activated". This is probably the reason for many of the peculiarities of the combustion of carbon

monoxide, including the " afterburning " and latent energy of combustion of this gas; this subject is referred to again in § 5 under *The lifetime of activated molecules*.

Kondratjewa (E. and H.) and Kondratjew (1936–38) have made quantitative measurements of the light yield of the characteristic CO-flame radiation from flames burning at various pressures and in the presence of moisture and of nitrogen. The light yield shows a maximum, per molecule of CO_2 formed, at a pressure of around 40 mm. Hg, the falling-off of the yield at high pressure being attributed to a quenching of the excited molecules. The light yield is reduced by the presence of nitrogen or water, the effect of the latter being so marked that Kondratjew attributed the decrease to a change in the mechanism of the combustion. With the dry gas at 100 mm. Hg pressure, approximate calculations indicate that one excited CO_2 molecule emits radiation for every 125 CO_2 molecules formed and, hence, that the excitation of the CO_2 molecules is due to the chemical process and is not thermal. This is supported by the observed excitation of the mercury line at 2537 A., when mercury vapour is introduced into the flame (Kondratjew, 1935), and the intervention of nitrogen at high pressures is also explained by reaction of normal N_2 molecules with the excited CO_2 molecules to form N_2 in the metastable $^3\Sigma$ state, from which oxides of nitrogen are formed by subsequent processes.

Explosion and engine spectra

The spectra of the explosions of mixtures of hydrocarbons, such as methane, acetylene, ethylene and propane, with air or oxygen have been studied by several investigators (Lauer, 1933; Bone and Outridge, 1936; Bell, 1937; Smith, 1940).

In general, the spectra of weak mixtures differ little from those of the inner cones of the ordinary flames, showing chiefly bands due to C_2, OH and CH, the weaker bands of CH at 3143 A. and 3628 A. appearing rather more easily. With mixtures giving more violent explosions the spectrum characteristic of the carbon monoxide flame is dominant, and for over-rich mixtures the explosion flame is luminous on account of carbon deposition. The cause of the carbon deposition has been discussed by Smith (1940) in relation to the cold flames and the occurrence of the Swan bands of C_2; it seems that the strong appearance of the CO-flame spectrum in the more violent explosions, especially in closed vessels, is associated with the afterburning of this gas.

Explosions of gasoline in engines have been studied by Clark and his colleagues (Clark and Thee, 1926; Clark and Henne, 1927; Clark and Smith, 1929, 30) and by Withrow and Rassweiler (1931, 32), who have examined the variation of spectrum with crank angle. In non-knocking explosions the spectra of the initial flame shows the bands of C_2, CH and OH, while the afterburning shows only the characteristic CO-flame spectrum, indicating that the afterburning is due to the relatively slow burning of the carbon monoxide formed.

Under knocking conditions the spectrum is usually continuous. Withrow and Rassweiler have also studied the effect on the spectrum of lead tetraethyl and other " dopes ". They find that removal of the knocking by these dopes changes the spectrum so that it again shows bands of C_2 and CH, and that at the same time lines of Pb and bands of PbO appear, showing that the lead tetraethyl is decomposed and that it is probably the decomposition products such as PbO which are effective in reducing the knock. The mechanism of the combustion in the engines under knocking and non-knocking conditions is discussed in relation to the radicals which are known spectroscopically to be present.

In addition to the above rather theoretical applications to the study of combustion in engines, spectroscopic methods are sometimes of use empirically for determining the type of combustion occurring. Thus the spectrum of an engine exhaust flame will sometimes distinguish between afterburning, which shows the CO flame spectrum, and burning of hydrocarbons and partly oxidized products, which shows the bands of CH and the hydrocarbon flame bands.

Other emission spectra

Among other flame spectra which have been examined, with a view to ascertaining what radicals and intermediary compounds are formed during the combustion, may be mentioned the oxy-ammonia flame (Rimmer, 1923), which shows strong bands at 3360 and 3370 A., due to the radical NH, a complex many-line structure in the visible region attributed to NH_2, the OH bands and the NO γ bands; the carbon disulphide flame (Emeléus, 1926; Kondratjew, 1930; Fowler and Vaidya, 1931), which shows bands due to S_2 and SO; and the flames of sulphur (Fowler and Vaidya, 1931), hydrogen disulphide (Fowler and Vaidya, 1931), phosphorus (Emeléus and Downey, 1934; Emeléus and Purcell, 1927; Emeléus, 1925), phosphine (Emeléus, 1927 b), arsenic (Emeléus, 1927, 29 b) and selenium (Emeléus and Riley, 1933). The cyanogen flame shows the well-known CN bands strongly, and in the presence of moisture the NH bands also appear (Fowler and Gregory, 1919). The CN violet bands, which have strong heads at 4216, 3883 and 3590 A. degraded to the ultra-violet, seem to occur in all flames which contain combined nitrogen and free carbon. Thus Barratt (1920) has observed the bands when organic compounds containing nitrogen are introduced into the coal-gas flame, in the coal-gas—nitrous-oxide flame, and in flames of carbon monoxide mixed with ammonia, but not in the carbon-monoxide—nitrous-oxide flame, the spectrum of which is similar to the carbon-monoxide—air flame. The hydrogen—nitrous-oxide flame (Fowler and Gregory, 1919; Fowler and Badami, 1931) is similar to the ammonia flame, showing bands due to NH, NH_2, OH and NO.

Bands which are probably due to BrO and IO have been observed by Vaidya (1937, 38) in flames of ethyl bromide and methyl iodide respectively, and Urey and Bates (1929) have reported bands due to ClO when chlorine is introduced into a hydrogen flame. The observation of these halogen monoxides in flames

may ultimately influence the theory of chain-breaking by inhibitors, most of which are halogen compounds.

Finch and Thompson (1930) have studied combustion processes in the cathode region of a direct-current electric discharge through combustible mixtures at low pressure, and, in an investigation of the spectrum of the discharge, were unable to detect the band spectrum of CO^+; hence they concluded that the discharge did not ionize the carbon monoxide.

§ 3. ABSORPTION SPECTRA

Some radicals and intermediary compounds which are formed at various stages during combustion processes can be detected, and even estimated quantitatively, by their absorption spectra. The advantages and limitations of this method are discussed by Oldenberg (1935). For the detection of absorption bands which have discrete rotational structure it is necessary to use spectrographs with very high resolving power and a fine slit, as the natural width of the absorption lines is so small; thus Oldenberg (1934) observed the absorption bands of OH with ease on a 21-ft. grating spectrograph, although previous workers, using smaller instruments, had failed; bands of CH, although occurring readily in emission in flames, are not usually detectable in absorption with instruments of moderate dispersion. Compounds whose absorption spectra are continuous, such as hydrogen peroxide, can be observed with instruments of small resolving power, but the absorption is usually less intense and the sensitivity of the method is correspondingly reduced. Compounds such as formaldehyde and benzene, whose spectra consist of bands with close "piled-up" rotational structure, show strong absorption on instruments of moderate resolving power, and the spectroscopic method is very suitable for their estimation. It may be remarked here that the region between 2300 and 1900 A., which is readily accessible on many small quartz spectrographs, using special photographic plates sensitized to the ultra-violet, is deserving of more attention than it has so far received, on account of the many organic compounds which absorb strongly in this region.

The OH radical

The OH band at 3064 A. is readily observed in absorption with instruments of high resolving power or by the spectrum-line reversal method. The bands have been observed in steam at temperatures above 1200° C. (Bonhoeffer and Reichardt, 1928), in discharges through water vapour and in flames.

Kondratjew and his colleagues (Kondratjew and Ziskin, 1936, 37 a; Avramenko and Kondratjew, 1937) have made quantitative measurements of OH absorption in flames and conclude that the concentration of hydroxyl is 1000 times that which would be produced by thermal dissociation of water at the temperature of the flame, indicating that the radicals are of chemical origin.

Oldenberg (1935) and Frost and Oldenberg (1936) have made an important contribution to the kinetics of the reactions of OH radicals by a quantitative

study of the variation of the absorption spectrum with time after the passage of an electric discharge through water vapour and through hydrogen peroxide. In discharges through water vapour the OH radicals are found to persist for as long as 0·4 sec., indicating that the destruction of the radicals is a three-body collision process, probably

$$OH + H + M = H_2O + M,$$

where M represents a third body. However, Kondratjew (1938) favours the reaction

$$OH + OH + M = H_2O_2 + M,$$

followed by decomposition of the hydrogen peroxide. Oldenberg points out that the rate of decay of the concentration of the radicals (as indicated by their absorption bands using intermittent electric discharges) shows that the persistence of the radicals is due to their long life and not to their steady formation from O and H atoms; this, he says, is supported by the emission of the OH bands in the fluorescence of steam sensitized with mercury to the Hg line at 2537 A., which could only occur as a two-stage process with the maintenance of an appreciable concentration of OH. For discharges through hydrogen peroxide, however, the OH radicals persist for less than 0·01 sec., indicating a bimolecular reaction between OH and H_2O_2.

Hydrogen peroxide

Hydrogen peroxide shows continuous absorption in the ultra-violet which becomes stronger to shorter wave-lengths (Urey, Dawsey and Rice, 1929). The absorption is relatively weak and the spectroscopic method cannot usually be regarded as a sensitive test, but Frost and Oldenberg (1936), by specially refined methods of measuring the light intensity, have been able to detect the peroxide at pressures as low as 0·01 mm. of mercury in a 150-cm. column, and have shown that hydrogen peroxide formed in discharges through water vapour is less than this amount. It must, therefore, be concluded that the formation of the peroxide, which is usually observed in such discharges, occurs in the collecting trap and not in the gas phase.

Slow combustion of hydrocarbons

The absorption spectrum of the intermediary products of the slow combustion of hydrocarbons has been studied by Egerton and Pidgeon (1933). The spectrum shows characteristic bands in the near ultra-violet due to formaldehyde, a continuous band of unknown origin around 2600 A., and a cut-off in the far ultra-violet due to organic acids. The band at 2600 A., which occurs only for the higher hydrocarbons (above butane), appears in the early stages of the combustion before the formaldehyde bands, and is not observed in the slow combustion of the aldehydes themselves. Ubbelohde (1935 b) has extended these experiments and has recorded the absorption spectra of a large number of organic compounds in an attempt to determine the nature of the intermediary compound responsible for the band at 2600 A.

In the slow combustion of propane, Newitt and Baxt (1939) have shown, by following the absorption spectrum with time, that propaldehyde is the only aldehyde present in the early stages of the oxidation, and that the lower aldehydes (acetaldehyde and formaldehyde) are formed later, the concentration of formaldehyde increasing with time. In this case the spectroscopic method has the advantage that it distinguishes between acetaldehyde and propaldehyde, this distinction being difficult to obtain by chemical tests; also, it is possible to follow the progress of the reaction spectroscopically without the interruptions to the reaction itself caused by chemical sampling.

Withrow and Rassweiler (1933, 34) found that in the early stages of combustion in engines the absorption spectrum shows the formaldehyde bands when the engine is knocking or running under conditions of incipient knock, this production of formaldehyde being so characteristic that it may be taken as a measure of the tendency of the fuel to knock. The formation of the aldehyde occurs prior to the burning of the charge. Non-knocking explosions do not usually show the formaldehyde absorption bands, but the addition of pro-knocks such as isopropyl nitrite or diethyl peroxide causes these bands to appear. Most substances which suppress knock in the engine also remove the formaldehyde bands from the spectrum, but this does not appear to be true of lead tetraethyl.

Nitrogen compounds in explosions

Bone and Newitt (1927) have examined, by self absorption from the flame itself, the absorption spectra of the flames and explosions of carbon monoxide and air, of carbon monoxide and nitrous oxide, and of mixtures of carbon monoxide with nitrous and nitric oxides. They found that the nitric-oxide formation, which is known chemically to occur strongly in such explosions at high pressures, did not show itself by its absorption spectrum in the initial explosion flame, although it was, of course, observed in absorption by the products of the combustion. This apparently indicates that the nitric-oxide formation is a secondary phenomenon dependent on a previous activation of the nitrogen. With the $CO+N_2O+NO$ explosions a characteristic banded absorption spectrum, with the strongest bands at 3680, 3545, and 3418 A., is observed. The origin of these bands is still in doubt (Newitt and Outridge, 1938; Thompson, 1939), but may be due to the radical $-NO_2$, or to HNO_2.

Other absorption spectra

Efforts to detect the absorption spectra of short-lived radicals in flames with a view to estimating their abundance and life have hitherto met with little success. Recent observations, however, of the absorption spectrum of CN in discharge tubes (White, 1938, 40), of NH in the thermal decomposition of ammonia (Franck and Reichardt, 1936), and of the Swan bands of C_2 in the thermal decomposition of carbon suboxide (Klemenc, Wechberg and Wagner, 1934 a, 34 b) at 200° C., show that the method has possibilities provided that spectrographs of adequate resolving power are employed.

§ 4. INFRA-RED EMISSION

Emission from flames

The infra-red spectra of flames were first investigated in detail by Julius in 1890 and by Paschen in 1893. The ordinary Bunsen-burner flame shows a very strong emission band at about 4·4 μ and a rather wider, but less intense, band near 2·8 μ. The former is characteristic of burning carbon monoxide; the band at 2·8 μ was at first attributed entirely to burning hydrogen, but is now known to be due to a blend of two bands, one characteristic of burning hydrogen and the other of burning carbon monoxide.

Comparison of these bands with the infra-red absorption of carbon dioxide and water (by Rubens and Aschkinass in 1898 and by Coblentz in 1905) shows that the band at 4·4 μ is due to CO_2 molecules (the absorption actually lies to slightly shorter wave-lengths, around 4·25 μ), while the broader band at 2·8 μ is characteristic of water vapour, but also falls on a strong combination band in the infra-red spectrum of carbon dioxide. The other fundamental absorption band of CO_2 at 14·9 μ which occurs strongly in absorption has not been observed in emission from the flame (Garner and Johnson, 1927; Bailey and Lih, 1929 a); this has been attributed by Gaydon (1941 a) to the comparatively long radiative life of the molecules in this state of vibration—so long that they lose energy by collision rather than by emission of radiation.

Bailey and Lih (1929 a) have studied the spectra of the flames of carbon monoxide, methane, hydrogen and coal-gas. In addition to the strong bands at 4·4 and 2·8 μ, they record a number of weaker bands, bands at 0·95, 1·45, 5·3, 5·5 and 6·7 being apparently associated with the burning of the hydrogen, while a band at 1·99 μ occurs only when carbon monoxide is being burnt, and is identified with a weak combination band in the spectrum of CO_2. The bands vary slightly in wave-length according to the temperature, and possibly other properties of the flame. This variation has been noted and commented on by Bailey and Lih and others; it is probably partly a genuine temperature effect, due to change in the rotational temperature distribution and possibly superposition of bands all corresponding to a similar change of vibrational quantum number, but starting from initially higher vibrational levels in the hotter flames, but is also probably largely affected by self-absorption. Flames of gases burning in nitrous oxide have also been studied by Bailey and Lih (1929 b); the hydrogen—nitrous-oxide flame shows characteristic bands, probably due to N_2O.

In a study of the spectra of the separated cones of coal-gas flames, Hartley (1932–33), found that the band at 4·4 μ is, relatively to the 2·8 μ band, stronger in the outer than in the inner cone; this agrees with the results of chemical sampling, which indicates that the formation of carbon dioxide occurs largely in the outer cone.

Measurements of the total radiation from various gases, per litre of gas burnt, were made by Helmholtz, Callendar, David and others (see Hartley, 1932–33) and it was found that the infra-red radiation from a carbon-monoxide

flame (presumably undried) was some 2·4 times that from a hydrogen flame, and that for hydrocarbons and coal-gas the amount of radiant energy emitted was the sum of that radiated by burning corresponding amounts of carbon monoxide and hydrogen. For a coal-gas flame the total amount of energy radiated, nearly all of which lies below 0·1 μ, is usually between 10 and 16 per cent of the total heat of the combustion, varying somewhat with conditions of air supply and burner design. For non-luminous flames the radiation is mostly due to the CO_2 molecules formed, while for luminous flames the emission is largely continuous, being due to emission from hot carbon aggregates, and in this case the proportion of heat radiated may reach 17 per cent.

Carbon-monoxide explosion-flames—Effect of catalysts

The effect of moisture, hydrogen and other catalysts on the infra-red emission from explosion flames of carbon-monoxide—oxygen mixtures has been the subject of a series of investigations by Garner and his colleagues (Garner, Johnson and Saunders, 1926; Garner and Johnson, 1927, 28). The emission is found to be markedly greater for the dry than for the wet gases; the difference is due to decreased radiation, in the presence of water, of the bands at 4·4 and 2·8 μ, the latter showing the greater decrease. Certain other catalysts, such as ethyl iodide and ethyl nitrate, have an even greater effect than water, the radiation being, in some cases, reduced to only 1 per cent of that emitted by the pure dry mixture. Carbon tetrachloride and nitrogen peroxide, however, act in the opposite manner, increasing the infra-red radiation slightly; this may be due to these substances removing traces of moisture. Careful drying of the gases also produces a big reduction in the flame speed, and this reduction is acribed by Garner and Johnson to the loss of energy by radiation from the flame.

The effect of hydrogen in reducing the infra-red radiation is also very marked (Garner and Roffey, 1929; Garner and Hall, 1930; Garner, Hall and Harvey, 1931; Bawn and Garner, 1932; Garner and Pollard, 1935). There is a definite step in the curve between the amount of infra-red radiation and the hydrogen concentration at 0·03 per cent of hydrogen. The infra-red radiation at 4·4 μ and 2·8 μ is known to be from vibrationally excited molecules of CO_2; Garner, Hall and Harvey conclude that during the combustion the CO_2 molecules contain a large fraction of the energy of the combustion. This forms at least 24 per cent of the total energy set free. These molecules are very stable towards collisions, but can lose their energy either by radiation or by resonance to other CO_2 molecules. As a result of resonance the activated molecules possess an apparently long duration of life.

Engine and explosion spectra

Steele (1931, 35) studied the infra-red emission from engines at various crank angles by a stroboscopic method, using filters which cut out certain regions of the infra-red spectrum, so that the variation of radiation in the bands at 4·4

and 2·8 μ could be studied with time. The emission continued to rise for some 20° of crank angle after the attainment of maximum pressure, and this was taken as evidence of the formation of metastable molecules of H_2O and CO_2.

David (1920, 26) also studied the variation of infra-red radiation with time during explosions, and suggested that the results indicated that the energy acquired by the freshly formed molecules of CO_2 and H_2O was not equally partitioned over the various internal degrees of freedom.

Radiation from gas fires

The character and quantity of the radiant energy emitted by gas fires has been of considerable interest to the gas industry, and the subject has been fully discussed by Hartley (1932–33), who examined the infra-red spectrum of the gas flame and of the fire itself. Some 90 per cent of the energy radiated by the fire is emitted by the radiants, brick and framework, while the characteristic flame radiation is reduced to about a third of that of the freely burning flame. The emission spectra of rods, coated with such materials as powdered silica, alumina, zirconia and uranium oxide and heated in a flame, have been measured quantitatively. The total radiation from fires whose radiants have been coated with such materials has also been measured.

§ 5. ENERGY CONSIDERATIONS

In this section, which might alternatively have been entitled "Miscellaneous", are included some examples of those many rather indirect applications of spectroscopy to combustion which provide the quantitative side to the kinetics of the reactions. These indirect applications have probably contributed far more to the theory of combustion than have the direct observations of flame and explosion spectra. The work is so closely linked with chemical kinetics generally, rather than to specific combustion problems, that it is only possible, within the scope of this *Report*, to include a few of the more striking applications.

Dissociation energies

An exact knowledge of the heats of formation or dissociation of the various compounds, initial, intermediary and final, which are present in any combustion process, is very necessary for a full understanding of the possible elementary reactions occurring. For chemically stable products, the heats of formation can usually be determined by direct measurement, but for the dissociation energies of diatomic molecules such as O_2, and the energy of formation of radicals such as OH, information derived from measurements of the band spectrum of the molecule concerned has to be used. The methods used to calculate the heats of dissociation of diatomic molecules have been conveniently summarized by Jevons (1932) and more recently by Herzberg (1939). The calculation of the dissociation energy may be resolved into two problems, the location of a dissociation limit and the determination of the dissociation products. The first problem, the location of the limits, is most satisfactorily solved when an absorption

band system can be followed up to the continuum which corresponds to dissociation, and such limits have been observed in the spectra of O_2, H_2 and the halogens. When the convergence of bands to a limit is not directly observable, it is possible to extrapolate to this limit from the vibrational energy levels which are known spectroscopically, and so obtain a close approximation to the dissociation energy (Birge and Sponer, 1926); the accuracy of the value depends on the number of vibrational energy levels which are known and which can be used in making the extrapolation. Information about the energy of dissociation can also be obtained sometimes from observation of an absorption continuum, of a predissociation, or by the excitation of atomic fluorescence. The dissociation products can usually be determined by consideration of the electronic types of the molecular levels (found from the rotational analysis of the spectrum) and of the line spectra of the constituent atoms, taking into account the Wigner-Witmer correlation rules.

Among the diatomic molecules whose energy of dissociation is known accurately from spectroscopic data, and which are of interest in combustion, are:

C_2 3·6, CH 3·47, H_2 4·4776, N_2 7·384, NO 5·29, O_2 5·082 and OH 4·3;

the values of the dissociation energy are expressed in electron volts. The values are also known for the halogens, the halogen hydrides, and many other diatomic molecules.

One of the outstanding needs at present is a determination of the dissociation energy of CO. Several attempts have been made to evaluate this (Gerö, 1936; Gerö, Herzberg and Schmid, 1937; Schmid and Gerö, 1937; Herzberg, 1937, 39; White, 1938–40 b), but although a dissociation limit is known, there is still some doubt about the dissociation products. Penney and Kynch, from calculations of the absorption spectra of polyenes, favour a value of 11·05 electron volts, in agreement with White. Knowledge of this energy is especially important as it would carry with it a knowledge of the heat of sublimation of carbon, a quantity which frequently occurs as an unknown in chemical equations relating to combustion processes.

Specific heats

For the calculation of theoretical flame and explosion temperatures and for the drawing up of accurate heat-balance equations to examine the efficiency of combustion, accurate values of the specific heats of all the combustion products and inactive diluents must be known. The direct measurement of specific heats at very high temperatures is difficult and subject to errors of some magnitude. The Boltzmann equipartition law enables the specific heat and other thermodynamic quantities of a gas to be calculated very accurately from spectroscopic data. The specific heat C_p may be considered as built up of three parts, due to the translation of the molecules (this being equal to the gas constant $\boldsymbol{R}$) to the rotational energy of the molecules (which may be regarded as a single fully

excited degree of freedom for diatomic gases and as two or three degrees of freedom for polyatomic gases, according to whether the molecule is linear or not) and to the vibrational energy (which becomes of increasing importance at high temperature). The calculation of this vibrational component requires values for the vibrational frequencies, which may be derived from infra-red or Raman spectra, or sometimes from the analysis of an electronic band spectrum which is known to involve the lowest state. In a few cases, of which the $^1\Delta$ level of O_2 is the best example, it is also necessary to take into account any electronic levels of particularly low energy. The use of such theoretical values of the specific heats, in preference to the rather doubtful experimental values, has been advocated by David (1934 a) and is now fairly general. The writer of this *Report* has not combed the literature for the numerous calculations of thermodynamic data which have been made; reviews summarizing some of the work up to 1934 have been given by Zeise (1933, 34); calculations directly designed for use in combustion work have also been made by Zeise (1938, 39); other papers giving values for the specific heats and other thermodynamic quantities at high temperatures for O_2, N_2, H_2, CO, CO_2, H_2O, etc., have been noted (Gordon and Barnes, 1933; Johnston and Walker, 1933, 35; Gordon, 1934; Johnston and Davis, 1934; Kassel, 1934).

Temperature measurements

The measurement of the temperature of flames is among the least satisfactory of physical measurements. The temperature of a reacting gas mixture is indeed rather difficult to define as the translational, rotational and vibrational temperatures of the molecules, and the temperature recorded by radiation methods may differ when the energy is not equipartitioned among the several possible forms. Spectroscopic methods have been applied to the measurement of flame temperatures in some cases.

The spectrum-line-reversal method has been used by Griffiths and Awbery (1929) and by Lewis and his colleagues (Jones, Lewis, Friauf and Perrott, 1931). David (1934 b, 37) has, however, criticized the method, and even Lewis says that a colour correction is necessary, that the lithium red line gives less satisfactory results than the sodium line, and that with thallium no true reversal of the absorption line can be obtained. Apart from the possibility that the sodium atoms are not in thermal equilibrium with the gases of the flame, it seems that the method must tend to give a value for the temperature of the outer sheath of the flame rather than of the body of the flame, although Griffiths and Awbery's comparison of the temperatures obtained from the reversal of the sodium line with values obtained by a hot-wire method do not indicate this.

It is also possible, in theory, to obtain flame temperatures from observation of the rotational intensity distribution of a band system which is emitted by the flame, such as the bands of CH or OH, but such observations as have been

made usually indicate abnormal temperatures due to lack of equipartition of the energy (Jevons, 1932; Herzberg, 1939), so that only a value for an effective rotational temperature of the molecules is obtained.

The lifetime of activated molecules

The radiation emitted by a flame or explosion depends not only on the nature and number of the various molecules, radicals, etc., which are excited, but also on the after-history of such excited molecules. An excited molecule may lose its energy either by radiation or on collision, and the way in which the activated molecules do actually lose their energy will depend on the radiative life of the molecules (i.e., the average time before which an undisturbed molecule would radiate freely) and on the collision life of the molecule. This after-history of the excited molecules will affect not only the character and amount of the light emitted during the combustion, but also the reactions occurring in the later stages of the combustion. Activated molecules are known to act as chain carriers, and perhaps in some cases as chain breakers. Lack of equipartition of the energy set free may modify the activation energy necessary for the continuation of the combustion processes.

The radiative life of an excited molecule can be calculated if quantitative measurements of the absorption spectrum corresponding to the excitation are available. Expressions, based on Einstein's derivation of the Planck radiation law, connecting the transition probability (reciprocal of the radiative life) with the absorption by a spectrum line, have been derived (Tolman, 1924; Korff and Breit, 1932). It is known generally that for permitted electronic transitions the radiative life is of the order 10^{-8} sec., while at atmospheric pressure the time between collisions is of the order 10^{-10} sec., although the time before energy is transferred to another form by collision (the collision life) may be much longer. Actual calculations of the radiative lifetimes of electronically excited molecules have been made in only a very few cases, largely owing to the difficulty in obtaining the very high resolving power necessary for the examination of the absorption spectrum. The best example is the calculation by Oldenberg and Rieke (1938) of the life of electronically excited OH radicals, which is found to be $3{\cdot}8 \times 10^{-6}$ sec. White (1940) has made similar calculations for the CN radical, using the band at 3883 A., and obtained, for a particular line, an f value (the number of dispersion electrons per molecule) of 0·1, which corresponds apparently to a radiative life of the order 10^{-8} sec.

Gaydon (1941 a) has made approximate calculations of the radiative life for the three fundamental vibrations of the CO_2 molecule; the radiative life of the asymmetrical vibration is found to be of the order 10^{-2} sec.; the transverse vibration has a life not exceeding 10^{-5} sec.; the symmetrical vibration, being inactive in the infra-red, will have a very much longer radiative life. These values are compared with the collision lives as derived from supersonic measurements, which show that molecules whose transverse vibration is excited are able to survive some 50,000 collisions with other CO_2

molecules before their energy is transferred into other forms, but that every thirtieth collision with a water molecule is effective. The conclusion is reached that, in the combustion of dry carbon monoxide, the CO_2 molecules formed may retain their vibrational energy for a few tenths of a second, but that the presence of water vapour will quickly dissipate the energy. The results are of importance in relation to the afterburning, and also give an explanation of the marked catalytic action of water in the combustion of CO and of the effect of moisture in reducing the infra-red radiation from the flame.

Spectroscopic evidence for the persistence of vibration has also been obtained in other cases. White (1938–40) notes that the absorption spectrum of CN radicals changes with time in such a way as to indicate that the vibrational energy persists for some 0·003 sec., corresponding to survival through some 10,000 collisions. Dwyer (1939) has shown that the vibrational energy of I_2 molecules produced by an electric discharge persists for 1/30 sec., corresponding to about 7000 collisions. Such resistance to loss of energy on collision by vibrationally activated molecules produces delay in the equipartition of the energy, and may affect the combustion in two ways. The time given to the vibrationally activated molecules may, as in the case of burning dry carbon monoxide, enable an appreciable percentage of the heat of combustion to be radiated away so that the flame temperature falls below the simply calculated theoretical value, and the flame speed may be lower owing to there being less energy available for the propagation of the combustion process; in other cases delay in exciting the vibrations of inactive diluents may result in the measured temperature exceeding the theoretical value (Lewis and von Elbe, 1938).

Calculations based on the resonance between the internal vibrations of a polyatomic molecule indicate that transfer of the vibrational energy from one form of vibration to another may result in dissociation, and Rosen (1933) has shown, by making some simplifying assumptions, that the short life of the HO_2 molecule, the existence of which is often postulated in combustion mechanisms, may be due to dissociation through such internal resonance. Rice (1933) has commented on details of the calculations, but agrees in principle. It has also been suggested (Gaydon, 1941 a, 41 b) that a similar effect in CO_2 may result in some of the molecules formed in the combustion quickly dissociating before the energy is lost by collision, thus contributing to the afterburning. Estimates of the rate of disappearance of OH radicals (Oldenberg and Rieke, 1939) made from measurements of the absorption spectrum indicate the formation of collision complexes which have a lifetime much greater than the duration of bimolecular collisions between stable molecules.

Electronic states and chemical reactions

The ways in which a chemical reaction may proceed are governed by many factors, such as activation energies. One factor which still presents a fairly open, if rather difficult, field for the investigator is that of the effect of the changes in the electronic structure of the molecules during reactions.

Smith (1940) has pointed out that the reactions which actually occur between molecules and radicals present in hydrocarbon combustion may be determined, at least partly, by their electronic states. It is to be expected that $^3\Pi$ molecules of C_2 will react readily with molecules of O_2 in the ground state, but the persistence of the CH bands is in accordance with the theoretical expectation that these molecules, being in a $^2\Pi$ state, will not react readily with O_2 molecules in the lowest $^3\Sigma$ state.

Schmid (1936) has shown that CO_2 is built from (i.e. would dissociate to) oxygen and an excited carbon monoxide molecule (probably CO $a^3\Pi$) and not from normal carbon monoxide in the $^1\Sigma$ ground state. This appears to necessitate a rearrangement of the electrons of the CO_2 molecules formed during the combustion of carbon monoxide and must influence the combustion processes (Gaydon 1940). It has also been recognized (Van Vleck, 1934 ; Voge, 1936) that in quadrivalent carbon, such as saturated hydrocarbons and organic compounds generally, the electronic energy is determined by various excited states of the carbon atom, especially the ^5S state, and not by the ground ^3P state of carbon ; it seems not unlikely that this may influence the combustion of hydrocarbons.

REFERENCES

Avramenko, L. and Kondratjew, V., 1937. *Acta. Physicochim. U.R.S.S.* **7**, 567.
Bailey, C. R. and Lih, K. H., 1929 a. *Trans. Faraday Soc.* **25**, 29.
Bailey, C. R. and Lih, K. H., 1929 b. *J. Chem. Soc.* 51.
Barratt, S., 1920. *Proc. Roy. Soc.* A, **98**, 40.
Bawn, C. E. H. and Garner, W. E., 1932. *J. Chem. Soc.* 129.
Bell, J., 1937. *Proc. Roy. Soc.* A, **158**, 429.
Birge, R. T. and Sponer, H., 1926. *Phys. Rev.* **28**, 259.
Bone, W. A. and Newitt, D. M., 1927. *Proc. Roy. Soc.* A, **115**, 41.
Bone, W. A. and Outridge, L. E., 1936. *Proc. Roy. Soc.* A, **157**, 234.
Bonhoeffer, K. F. and Haber, F., 1928. *Z. phys. Chem.* A, **137**, 263.
Bonhoeffer, K. F. and Reichardt, H., 1928. *Z. Elektrochem.* **34**, 652.
Burch, G. J., 1886. *Nature, Lond.*, **35**, 165.
Clark, G. L. and Henne, A. L., 1927. *J. Soc. Auto. Eng.* **20**, 264.
Clark, G. L. and Smith, H. A., 1929. *J. Phys. Chem.* **33**, 659.
Clark, G. L. and Smith, H. A., 1930. *J. Phys. Chem.* **34**, 1924.
Clark, G. L. and Thee, W. C., 1926. *Industr. Engng. Chem.* **18**, 528.
David, W. T., 1920. *Phil. Mag.* **39**, 84.
David, W. T., 1926. *Phil. Mag.* **1**, 334.
David, W. T., 1934 a. *Engineer, Lond.*, **158**, 476.
David, W. T., 1934 b. *Engineering*, **138**, 475.
David, W. T., 1937. *Phil. Mag.* **23**, 251.
Dwyer, R. J., 1939. *J. Chem. Phys.* **7**, 40.
Egerton, A. C. and Pidgeon, L. M., 1933. *Proc. Roy. Soc.* A, **142**, 26.
Emeléus, H. J., 1925. *J. Chem. Soc.* 1362.
Emeléus, H. J., 1926. *J. Chem. Soc.* 2948.
Emeléus, H. J., 1927. *J. Chem. Soc.* 783.
Emeléus, H. J., 1929 a. *J. Chem. Soc.* 1733.
Emeleus, H. J., 1929 b. *J. Chem. Soc.* 1846.
Emeléus, H. J. and Downey, W. E., 1924. *J. Chem. Soc.* 2491.
Emeléus, H. J. and Purcell, R. H., 1927. *J. Chem. Soc.* 788.

EMELÉUS, H. J. and RILEY, H. L., 1933. *Proc. Roy. Soc.* A, **140**, 378.
FARKAS, L., 1931. *Von den Kohlen- und Mineralölen*, **4**, 35.
FINCH, G. I. and THOMPSON, H. H., 1930. *Proc. Roy. Soc.* A, **129**, 314.
FOWLER, A. and BADAMI, J. S., 1931. *Proc. Roy. Soc.* A, **133**, 325.
FOWLER, A. and GREGORY, C. C. L., 1919. *Philos. Trans.* A, **218**, 351.
FOWLER, A. and VAIDYA, W. M., 1931. *Proc. Roy. Soc.* A, **132**, 310.
FRANCK, H. H. and REICHARDT, H., 1936. *Naturwissenschaften*, **24**, 171.
FROST, A. A. and OLDENBERG, O., 1936. *J. Chem. Phys.* **4**, 642 and 781.
GARNER, W. E. and HALL, D. A., 1930. *J. Chem. Soc.* 2037.
GARNER, W. E., HALL, D. A. and HARVEY, F. E., 1931. *J. Chem. Soc.* 641.
GARNER, W. E. and JOHNSON, C. H., 1927. *Phil. Mag.* **3**, 97.
GARNER, W. E. and JOHNSON, C. H., 1928. *J. Chem. Soc.* 280.
GARNER, W. E., JOHNSON, C. H. and SAUNDERS, S. W., 1926. *Nature, Lond.*, **117**, 790.
GARNER, W. E. and POLLARD, F. H., 1935. *J. Chem. Soc.* 144.
GARNER, W. E. and ROFFEY, F., 1929. *J. Chem. Soc.* 1123.
GAYDON, A. G., 1940. *Proc. Roy. Soc.* A, **176**, 505.
GAYDON, A. G., 1941 a. *Proc. Roy. Soc.* A, **178**, 61.
GAYDON, A. G., 1941 b. *Nature, Lond.*, **148**, 226.
GAYDON, A. G., 1942. *Proc. Roy. Soc.* A, **179**, 439.
GAYDON, A. G. and PEARSE, R. W. B., 1935. *Proc. Roy. Soc.* A, **148**, 312.
GEIB, K. H. and VAIDYA, W. M., 1941. *Proc. Roy. Soc.* A, **178**, 351.
GERÖ, L., 1936. *Z. Phys.* **100**, 374.
GERÖ, L., HERZBERG, G. and SCHMID, R., 1937. *Phys. Rev.* **52**, 467.
GORDON, A. R., 1934. *J. Chem. Phys.* **2**, 65.
GORDON, A. R. and BARNES, C., 1933. *J. Chem. Phys.* **1**, 297.
GRENAT, H., 1931. *C.R. Acad. Sci., Paris*, **192**, 1553.
GRIFFITHS, E. and AWBERY, J. H., 1929. *Proc. Roy. Soc.* A, **123**, 401.
HARTLEY, H., 1932–33. *Trans. Inst. Gas Eng.* **82**, 466.
HEDFELD, K., 1931. *Z. Phys.* **68**, 610.
HERZBERG, G., 1937. *Chem. Rev.* **20**, 145.
HERZBERG, G., 1939. *Molecular Spectra and Molecular Structure*, **1** (New York : Prentice-Hall).
HORI, T., 1930. *Z. Phys.* **59**, 91.
HUND, F., 1926. *Z. Phys.* **36**, 657.
JACK, D., 1927. *Proc. Roy. Soc.* A, **115**, 373.
JEVONS, W., 1932. *Report on Band Spectra of Molecules* (London : The Physical Society).
JOHNSTON, H. L. and DAVIS, C. O., 1934. *J. Amer. Chem. Soc.* **56**, 271.
JOHNSTON, H. L. and WALKER, M. K., 1933. *J. Amer. Chem. Soc.* **55**, 177.
JOHNSTON, H. L. and WALKER, M. K., 1935. *J. Amer. Chem. Soc.* **57**, 682.
JONES, G. W., LEWIS, B., FRIAUF, J. B. and PERROTT, G. ST. J., 1931. *J. Amer. Chem. Soc.* **53**, 869.
KASSEL, L. S., 1934. *J. Amer. Chem. Soc.* **56**, 1838.
KITAGAWA, T., 1936. *Proc. Imp. Acad., Tokyo*, **12**, 281.
KLEMENC, A., WECHSBERG, R. and WAGNER, G., 1934 a. *Z. Elektrochem.* **40**, 488.
KLEMENC, A., WECHSBERG, R. and WAGNER, G., 1934 b. *Z. phys. Chem.* **170**, 97.
KONDRATJEW, V., 1930. *Z. Phys.* **63**, 322.
KONDRATJEW, V., 1935. *Acta Physicochim. U.R.S.S.* **2**, 126.
KONDRATJEW, V., 1936. *Acta Physicochim. U.R.S.S.* **4**, 556.
KONDRATJEW, V., 1938. *Acta Physicochim. U.R.S.S.* **8**, 315.
KONDRATJEW, V. and ZISKIN, M., 1936. *Acta Physicochim. U.R.S.S.* **5**, 301.
KONDRATJEW, V. and ZISKIN, M., 1937 a. *Acta Physicochim. U.R.S.S.* **6**, 307.
KONDRATJEW, V. and ZISKIN, M., 1937 b. *Acta Physicochim. U.R.S.S.* **7**, 65.
KONDRATJEWA, E. and KONDRATJEW, V., 1937. *Acta Physicochim. U.R.S.S.* **6**, 748.
KONDRATJEWA, E. and KONDRATJEW, V., 1938. *Acta Physicochim. U.R.S.S.* **8**, 481.

KONDRATJEWA, H. and KONDRATJEW, V., 1936. *Acta Physicochim. U.R.S.S.* **4**, 547.
KONDRATJEWA, H. and KONDRATJEW, V., 1937. *Acta Physicochim. U.R.S.S.* **6**, 625.
KORFF, A. and BREIT, G., 1932. *Rev. Mod. Phys.* **4**, 47.
KRATZER, A., 1924. *Z. Phys.* **23**, 298.
LAUER, F. J., 1933. *Z. Phys.* **82**, 179.
LAVIN, G. J. and STEWART, F. B., 1929. *Proc. Nat. Acad. Sci.* **15**, 829.
LEWIS, B. and VON ELBE, G., 1935. *Engineering*, **136**, 168.
LEWIS, B. and VON ELBE, G., 1938. *Combustion Flames and Explosions of Gases* (Cambridge).
MULLIKEN, R. S., 1927. *Phys. Rev.* **30**, 785.
MULLIKEN, R. S., 1930. *Rev. Mod. Phys.* **2**, 60.
MULLIKEN, R. S., 1931. *Rev. Mod. Phys.* **3**, 89.
MULLIKEN, R. S., 1932. *Rev. Mod. Phys.* **4**, 1.
NEWITT, D. M. and BAXT, L. M., 1939. *J. Chem. Soc.* 1711.
NEWITT, D. M. and OUTRIDGE, L. E., 1938. *J. Chem. Phys.* **6**, 753.
OLDENBERG, O., 1934. *Phys. Rev.* **46**, 210.
OLDENBERG, O., 1935. *J. Chem. Phys.* **3**, 266.
OLDENBERG, O. and RIEKE, F. F., 1938. *J. Chem. Phys.* **6**, 439 and 779.
OLDENBERG, O. and RIEKE, F. F., 1939. *J. Chem. Phys.* **7**, 485.
PRETTY, W. E., 1928. *Proc. Phys. Soc.* **40**, 71.
RASSWEILER, G. M. and WITHROW, L., 1931. *Industr. Engng. Chem.* **23**, 769.
RASSWEILER, G. M. and WITHROW, L., 1932. *Industr. Engng. Chem.* **24**, 528.
RICE, O. K., 1933. *J. Chem. Phys.* **1**, 625.
RIMMER, W. B., 1923. *Proc. Roy. Soc.* A, **103**, 696.
ROSEN, N., 1933. *J. Chem. Phys.* **1**, 319.
SCHMID, R., 1936. *Z. Phys.* **99**, 626.
SCHMID, R. and GERÖ, L., 1937. *Z. phys. Chem.* B, **36**, 105.
SMITH, E. C. W., 1940. *Proc. Roy. Soc.* A, **174**, 110.
SMITHELLS, A., 1901. *Phil. Mag.* **1**, 476.
STEELE, S., 1931. *Nature, Lond.*, **128**, 185.
STEELE, S., 1935. *Nature, Lond.*, **135**, 268.
THOMPSON, H. W., 1939. *J. Chem. Phys.* **7**, 136.
TOLMAN, R. C., 1924. *Phys. Rev.* **26**, 693.
UBBELOHDE, A. R., 1935 a. *Proc. Roy. Soc.* A, **152**, 354.
UBBELOHDE, A. R., 1935 b. *Proc. Roy. Soc.* A, **152**, 378.
UREY, H. C. and BATES, J. R., 1929. *Phys. Rev.* **34**, 1541.
UREY, H. C., DAWSEY, L. H. and RICE, F. O., 1929. *J. Amer. Chem. Soc.* **51**, 1371.
VAIDYA, W. M., 1934. *Proc. Roy. Soc.* A, **147**, 513.
VAIDYA, W. M., 1937. *Proc. Indian Acad. Sci., Bangalore*, **6A**, 122.
VAIDYA, W. M., 1938. *Proc. Indian Acad. Sci., Bangalore*, **7A**, 321.
VAIDYA, W. M., 1941. *Proc. Roy. Soc.* A, **178**, 351.
VAN VLECK, J. H., 1934. *J. Chem. Phys.* **2**, 20.
VOGE, H. H., 1936. *J. Chem. Phys.* **4**, 581.
VON ELBE, G. and LEWIS, B., 1937. *J. Amer. Chem. Soc.* **59**, 976.
WATSON, W. W., 1924. *Astrophys. J.* **60**, 145.
WESTON, F. R., 1925. *Proc. Roy. Soc.* A, **109**, 176 and 523.
WHITE, J. U., 1938. *J. Chem. Phys.* **6**, 294.
WHITE, J. U., 1940. *J. Chem. Phys.* **8**, 79 and 459.
WITHROW, L. and RASSWEILER, G. M., 1933. *Industr. Engng. Chem.* **25**, 923 and 1359.
WITHROW, L. and RASSWEILER, G. M., 1934. *Industr. Engng. Chem.* **26**, 1256.
ZEISE, H., 1933. *Z. Elektrochem.* **39**, 758 and 895.
ZEISE, H., 1934. *Z. Elektrochem.* **40**, 662 and 885.
ZEISE, H., 1938. *Feuerungstech.* **26**, 145.
ZEISE, H., 1939. *Z. Elektrochem.* **45**, 456.

FRICTION BETWEEN SOLID BODIES

By ROBERT SCHNURMANN

§1. INTRODUCTION

The resisting force of friction is called into action at the common boundary of two materials when a given tangential force is applied so that one of the materials either tends to slide, or actually slides, over the surface of the other. The phenomenon of friction is common to all substances, whether solid, liquid, or gaseous, which may be in contact with each other and subject to forces tending to cause relative motion, but the laws correlating the magnitude of the frictional forces with the forces which call them into existence and the state of motion produced by given tractive forces differ widely with the nature of the substances. This article will be restricted to a consideration of the friction between solid bodies, without which there could be no walking, no driving, no riding and no railway traffic.

The simplest representation of the bodies of the real physical world is in terms of the so-called rigid body, of which satisfactory use is made in treating many problems of statics and dynamics. In the first instance the surface of a rigid body may be postulated to be perfectly smooth, so that in mechanics the mutual action between two such bodies in contact can be regarded as a single force which is perpendicular to the plane of contact, and which may assume any magnitude. However, this simple picture is not consistent with the actual facts as they present themselves when, for instance, a solid body without external support maintains its position on an inclined plane independent of the angle α at which the plane is tilted to the horizontal, provided that α does not exceed a limiting value α_0, which is called the angle of repose. The weight of the body and the reaction of the inclined plane on which it is resting could not maintain equilibrium by themselves unless the mutual action between the body and the plane exhibited also a tangential component—the force of friction. Under static conditions this force may assume any magnitude between zero and the limiting value of the force of static friction, which can be measured either by the tilt method (measurement of the angle of repose), or by elastically restraining one friction element and applying a tangential force to the other one; the magnitude of the elastic force increases, but is balanced by the correspondingly growing force of static friction, until the elastic force reaches the limiting, that is, the greatest possible, value of the force of static friction, when static friction ceases to

balance the elastic force, and the elastically restrained element must commence to slide. The force of friction is always a function of the velocity of sliding and, depending upon the shape of the force-velocity characteristic of sliding of a particular combination of friction elements, the force of friction at a given velocity of sliding may be either equal to or different from the limiting value of the force of static friction.

§ 2. AMONTONS' LAW

Ever since the days of Leonardo da Vinci (*c.* 1500), it has been known that a slider would move on an inclined plane if the ratio between the tangential and normal components of the gravitational force acting at the area of contact exceeded one-fourth, the size of the contact area being immaterial. Amontons (1699) measured the force F required to move a solid body, with a macroscopically plane surface, relative to a similar horizontal surface, and he observed that for various materials the relation $F=\mu N$ held when the normal load N was varied within a limited range; the ratio $\mu=F/N$, which is called the coefficient of friction, was approximately independent both of the area of contact and of the physical properties of the solids, and was about one-third when sliding began. Numerous attempts have been made to use Amontons' law as a criterion for the mechanism of friction, and Coulomb's (1785) proposal, which was developed from de la Hire's (1699) hypothesis of the mechanism of friction, attracted much attention for a long time. It was an idea of de la Hire, and is a proved fact to-day, that even the macroscopically plane surface of a solid body is a maze of irregularities. According to de la Hire these irregularities should interlock, and either deform or abrade each other during relative motion. Coulomb visualized two possible causes of friction, a cohesion between the surfaces in contact, and the necessity of lifting the slider over the irregularities on the lower surface; cohesion would be proportional to the area of contact and independent of the load, whilst the force required for lifting the slider would be directly proportional to the load and independent of the area of contact, so that the second cause appeared to Coulomb to predominate (also see Bikerman, 1941), since Amontons' law $F=\mu N$ is approximately valid for unlubricated friction, at least within a limited range of normal loads. Bikerman and Rideal (1939) have suggested that in the light of present knowledge Coulomb's criterion requires reconsideration on account of the difference between the real area A and the apparent or nominal area A_0 of contact, since the hypothetical cohesion would, of course, be proportional to A, the real area of contact, and if A is proportional to the normal load N, as has been postulated by both Tomlinson (1939) and by Bowden and Tabor (1939), Amontons' empirical law would be consistent with both mechanisms. But, true as it is that surfaces exhibit roughnesses of at least molecular dimensions, and that friction must arise at the highest of these, the virtue of Amontons' empirical law is that it affords a criterion as to whether the relative motion between

two solids in frictional contact proceeds by sliding or by tearing,* without being concerned with the mechanism of sliding (Schnurmann, 1942); or, in the words of Sir William Hardy and Miss Doubleday (1922), that Amontons' law " holds so long as the surfaces, or rather the whole system, solid surfaces and lubricant, remain unchanged ", so that " cases in which the traction alters the texture of the solid " are outside its range. The ideas of *regular* interlocking between two " smooth " surfaces (Merchant, 1940) and of *direct* contact between them in air belong to bygone days (Schnurmann, 1940 a).

§ 3. THE ACTUAL AREA OF CONTACT

The smoothest physical surface is the cleavage plane of a single crystal. The study of the specular reflection of molecular rays of helium of certain wave-lengths (Estermann and Stern, 1930) from the cleavage plane of a single crystal of lithium fluoride, for instance, showed that the intensity of the reflected beam in a high vacuum decreased steeply when the angle of incidence between the primary beam and the cleavage plane exceeded about 20°, and it was calculated from this result that the cleavage plane exhibited irregularities of height of about two wave-lengths of the incident beam (about 1 to 2×10^{-8} cm.). The plausible assumption was made that these molecular irregularities on a cleavage plane were due to the thermal movement of the ions of the crystal lattice (LiF), and this was supported by the experimental evidence that the reflecting power of the cleavage plane of the crystal increased when its temperature was lowered. Highly polished surfaces would appear very rough by these standards, the height of their irregularities being of the order of several hundred molecular diameters. Since light has a much larger wave-length than molecular rays, it is a considerably less sensitive tool for the inspection of surface finish. A surface which is " optically flat " exhibits, in fact, irregularities of a height of almost 10^{-5} cm., so that molecular rays which strike it at a glancing angle of more than about 10^{-3} radians are scattered from it but not specularly reflected (Schnurmann, 1938). The highest of these irregularities carry the load when one solid is put upon the surface of another, so that there is no regular interlocking over the total nominal area of contact, and it is easy to demonstrate that the number of " legs " with which a polished metal surface stands upon a fused glass plate is large; their integrated area is, however, only a small fraction of the nominal area of contact when the normal load is insufficient to exceed the elastic limit of either the metal or the glass. For instance, the polished surface of a copper block of 673 g. weight was made to slide slowly over the " clean " surface of a fused glass plate. Amontons' law applied to the conditions of this experiment, and observation with the naked eye did not reveal any damage to either surface, while microscopic inspection of the metal surface (magnification about 330 times) showed a few parallel lines on it (figure 1). A latent image of the actual area of

* Coulomb discontinued his measurements, which confirmed Amontons' results, whenever, scratching of the surfaces became apparent.

contact could be developed when the glass plate was treated with a solution of silver nitrate to which a reducing agent had been added. The developed " traces " had the appearance of very thin scratches (Schnurmann, 1940 e, 42), and could be resolved under the microscope into multiplets of predominantly parallel lines which take up a very small proportion of the nominal area of contact (figure 2). Both the latent image and its developed " trace " could be removed by acid cleaning.

§ 4. THE ELEMENTARY MECHANISM OF FRICTION

The actual area of contact, in the experiment just described, was not an area of *direct* contact between copper and glass, because free surfaces always carry matter which they have adsorbed from the surrounding atmosphere. If the atmosphere is humid and also contains dust and greasy matter in suspension, as air does unless it has been cleaned carefully, foreign matter will condense and settle down in addition to any adsorbed layers. The surface of a solid in air can be made " mechanically clean " by freeing it from visible particles and from thick layers of condensed liquid, but it is impossible to remove, by *any* method of " cleaning ", the whole of the foreign matter so long as the surface remains exposed to the atmosphere. Even the drastic procedure of removing the contaminated surfaces completely by abrasion does not lead to a naked surface, since re-contamination from atmospheric sources begins instantaneously. The contaminating layers affect the " dry friction "* or " ordinary friction " behaviour of a given combination of solids; thus it is well known that the frictional resistance to slipping of locomotive wheels on the rails depends upon the hygrometric conditions of the air. Such inevitable layers of contamination are responsible for the wide variations in the values of coefficients of static " dry friction " which are obtained with different experimental methods and under varying experimental conditions. When the layers become very thin they may assume adhesive properties, and monomolecular layers exhibit a high degree of tenacity. The engineer applies lubrication to avoid such hazards in those rubbing machine parts between which a reduction of the frictional resistance is essential.

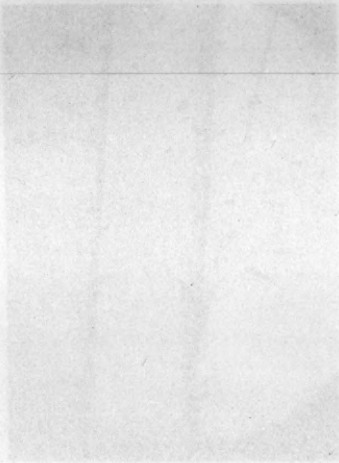
Lubricated friction under running conditions is a hydrodynamical problem. In a journal bearing, for instance, operating under light loads and high speeds, the only physical properties that matter are the geometrical dimensions of the

* In view of the widely differing terminology in papers on friction, the following definitions will be used in this article for the terms " solid friction ", " dry friction ", " ordinary friction", " boundary lubrication " and " friction elements ". " Solid friction " connotes the friction between naked surfaces. Both " dry " and " ordinary friction " arise in air when no lubricant has been added intentionally and the surfaces carry films of adsorbed and condensed matter. " Boundary lubrication " covers all those cases in which the solid bodies are separated by a film of insufficient thickness for hydrodynamic lubrication. " Friction elements " are the bodies which are in frictional contact.

Figure 1. " Traces " (magnification about 330 times) on the polished surface of a block of copper which was allowed to slide on an inclined glass plate in air.

Figure 2. Developed " traces " on a glass plate on which a copper block was allowed to slide in air. The inclination of the plate was 17·5°, and the weight of the copper block was 673 g.

To face page 74



bearing and the viscosity of the lubricant. This thick-film lubrication is a problem involving the viscosity of liquids and throws no light upon the friction of solid bodies. But even in cases in which a small friction resistance is desirable, and thick-film lubrication can be applied, disturbances occur. The thickness of the load-carrying film is reduced appreciably during starting and stopping periods and during shock, when the ensuing boundary lubrication (sometimes also called non-viscous or thin-film lubrication) shows features of similarity with " dry friction " and confronts the engineer with the question of the mechanism by which mechanical energy is converted into frictional heat.

We have seen that if two surfaces are in contact the friction must arise at those irregularities which carry the load. Suppose the surfaces were naked, freed from the last monomolecular adsorption layer and—in the case of metals —even from oxide layers; the irregularities could suffer both *elastic deformation* and *abrasion* during sliding, and a thermodynamical criterion can be applied for deciding which of these two mechanisms is predominant in solid friction (between such naked surfaces). The problem is, therefore, to determine whether over a wide range of temperatures, from the melting point to the neighbourhood of the absolute zero of temperature, the amount of mechanical energy converted into heat in a friction experiment varies with the temperature of the system.

Quite generally, whenever contact is established between two solids, adiabatic compression of a portion of the surface of a material is accompanied by a local temperature change which gives rise to a small flow of heat. If the conditions of the experiment were such that the deformed portion could return to its original configuration before any heat had been conducted away, the total amount of thermal energy that could be converted into mechanical energy would be recovered in the restoration phase. Actually, some heat flow, however minute, will occur during every cycle of compression and restoration, the thermal diffusivity of the material determining the rate of energy dissipation. Since both the adiabatic compression energy and the thermal conductivity exhibit a marked dependence upon temperature between the melting point and the absolute zero, the energy loss which is perceived as friction should exhibit a measurable temperature coefficient, if deformation of the surface irregularities is the elementary mechanism of solid friction. On the other hand, if abrasion of molecules (even monatomic) from the irregularities of the surfaces in contact were the predominant action in solid friction, virtual independence of temperature could be expected, since the energy loss incurred in detaching molecules from the surfaces would correspond to the latent heat of vaporization, which has only a very slight temperature coefficient below the melting point of a material.

§ 5. THE TEMPERATURE COEFFICIENT OF STATIC FRICTION

The amount of mechanical energy converted into heat in a friction experiment could be measured directly in a calorimeter at various temperatures, but if energy measurements are to be of any value in the investigation of the elementary mechanism of solid friction it is essential to prevent abraded matter from coalescing or from adhering to the surfaces in frictional contact. These difficulties are not encountered when the limiting value of the force of static friction is measured over a wide range of temperatures. If the force of friction were a function of the temperature alone, and independent of the velocity of sliding, the variation of static friction with temperature would indicate directly the

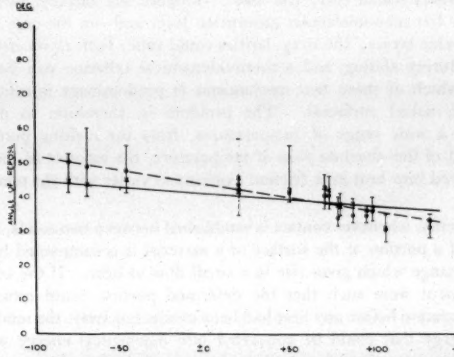


Figure 3. Angles of repose of naked cadmium at various temperatures in a high vacuum. The crosses and circles indicate the mean values of the critical angles obtained by tilting the friction chamber in opposite directions, whereas the range of individual values is represented by the vertical lines.

temperature coefficient of the energy dissipated, by solid friction, in a given displacement. There is reason to believe that this is the case for naked *metal* surfaces over a wide range of velocities of sliding. Experiments on cadmium cleaned by volatilization in a high vacuum at various temperatures at which the whole friction chamber was baked out indicate the existence of a small temperature coefficient (Schnurmann, 1941 a). A conservative analysis of the results of measurements between $-100°$ C. and $+100°$ C. (continuous line, figure 3) leads to a temperature coefficient of the angle of repose $\frac{d\bar{\alpha}}{dt} = -3{\cdot}5 \times 10^{-2}$ and of the coefficient of friction $\frac{d\mu}{dt} = -7{\cdot}5 \times 10^{-4}$. The dotted line in figure 3 shows that a value $\frac{d\bar{\alpha}}{dt} = -8 \times 10^{-2}$ is also consistent with the experimental

results. It might be thought that measurements with very soft materials like cadmium, in which deformation in the neighbourhood of points of contact is certain to occur (on account of the inevitable stress concentration there), would yield little information as to the nature of friction, since the plastic properties of cadmium depend upon temperature. However, no noticeable temperature coefficient of friction could be expected if abrasion alone were responsible for the elementary mechanism of friction. The fact that these properties and the coefficient of solid friction both depend upon temperature is evidence of the part played by deformation in the mechanism of solid friction.

If we assume that sliding proceeds by means of elastic deformation of the surface irregularities, adiabatic compression takes place and the energy associated with this process tends to zero as the absolute zero of temperature is approached. The process itself is irreversible because of the finite thermal conductivity of the materials, which, however, tends to zero in the neighbourhood of the absolute zero of temperature (see Borelius, 1935), so that the process tends to become reversible in the immediate neighbourhood of the absolute zero. In this case, entropy changes should vanish (Nernst's heat theorem), and, therefore, solid friction would be unable to assume any finite value. In addition, when the temperature of the surface of the lower-melting friction element reaches the melting point the force of friction between non-porous solids assumes a very small value, because then the molten surface layers act as a lubricant, as indicated by the low values observed with ice at temperatures near its melting point (Jacob, 1912; Bowden and Hughes, 1939 a). From the absolute zero of temperature to the melting point the sign of the temperature coefficient of solid friction would thus change at least once, although no prediction could be made about either the existence or the sign of a temperature coefficient in the intermediate range of temperatures. The above-mentioned experiments on cadmium exhibit a small value of the temperature coefficient; independently an experiment which involves deformation only should be made over a wide range of temperatures to determine any difference between the order of magnitude of the temperature coefficient in the case of a friction experiment and that of an experiment in which no action other than elastic deformation is involved. The rate of energy dissipation in the sensibly elastic range has been studied at room temperature when a solid sphere rebounds from the surface of a finite mass of elastic solid (Schnurmann, 1941 b). An extension of this experiment to lower temperatures should provide the desired information.

There is much uncertainty as to the effect of temperature upon " dry friction ". Jacob (1912) and Shaw and Leavey (1930) have attempted to measure the effect of temperature upon the static friction in a chamber in which the pressure was reduced to about 10^{-1} mm. Hg. The static friction appeared sometimes to increase and sometimes to decrease as the temperature rose. Miss Jacob obtained a minimum of the angle of repose of glass on glass at about 150° c., the coefficient of static friction being about one-half, or even as small as one-third,

of the value at room temperature. Shaw and Leavey obtained similar initial readings with silver on silver, copper on copper, and nickel on nickel. After repeated heating and cooling, however, the friction rose and finally settled to values which were in general practically constant between 15° c. and 350° c. Bowden and Hughes (1939 b) applied a deceleration method for the measurement of the path travelled by a small hollow metal cylinder over a lower cylindrical surface of smaller diameter. The former cylinder was projected along the lower surface by means of a hammer blow, the pressure in the friction chamber being maintained below 10^{-6} mm. Hg. and the temperature of the lower surface varied. The friction appeared sometimes to decrease in the neighbourhood of the melting point of the lower-melting friction element and sometimes to increase.

Interesting results were obtained by Bowden and Hughes (1939 a) when they made experiments on ice and snow in an ice cave at the Jungfraujoch in Switzerland. These results showed convincingly that frictional heating played a large part in melting a water film during sliding. It was observed that the coefficient of static friction of ice was about 0·5 between $-110°$ c. and $-71°$ c. and decreased upon approaching the melting point, being 0·3 at $-12°$ c. and 0·05–0·15 at 0° c.

§ 6. THE FORCE OF STATIC FRICTION

Attempts have been made in recent years to understand the nature of static friction forces, and a number of theories have been advanced (Frenkel, 1926; Prandtl, 1928; Tomlinson, 1929; Derjaguin, 1934) in which static friction is considered as the result of hypothetical conservative forces. Stevens (1899) made interferometric measurements from which he concluded that when a proper force was applied under conditions of " dry friction " there was a motion of one surface over another which took place before the body actually began to slide; this motion, unlike that produced when static friction is overcome, is imperceptible under ordinary conditions, and the surface returns to its first position when the force is removed; " these minute distances passed over by the plate are nearly (probably exactly) proportional to the force applied to produce the motion ". He observed that a displacement by one- or two-tenths of the wave-length of yellow light (i.e. by about 10^{-5} cm.) marked the limit of the transition to sliding friction. Rankin (1926, 28–29) undertook similar measurements using the oscillating valve ultra-micrometer method, and detected conservative forces in displacements up to 3×10^{-6} cm. Tomlinson observed the oscillations of a pendulum, the damping of which was provided by frictional contact, and arrived at the conclusion that the forces of interaction between the surfaces in frictional contact ceased to be conservative for displacements of less than 1×10^{-7} cm. The experimental difficulties with these methods are considerable, and the elastic forces apt to occur *inside* the friction elements may mask the actual forces of interaction between the surfaces in contact. Mandelstam and Haykin therefore proposed a dynamic method by means of which a piezo-electric quartz crystal was excited to perform rapid mechanical oscillations while

a small plate rested upon it. The oscillations of the crystal, which was gripped between two steel needles pressing against the centres of opposite faces perpendicular to the optical axis, were found suitable by Haykin, Lissovsky and Solomonovich (1939) for putting this method into practice. Piezo-electric quartz resonators possess a very sharp curve of resonance, the shape of which is highly sensitive even to slight disturbances introduced by the forces of interaction between the quartz crystal and the body resting upon it; thus the shape of the resonance curve, with regard to its dependence upon the amplitude of the oscillations, can be used to indicate the character of the interaction forces. Two piezo-electric quartz resonators, having frequencies of 70,000 and 150,000 cycles respectively, were used and it was found that the forces of interaction between the quartz crystal and the body resting upon it were conservative for small amplitudes of not more than about 4×10^{-6} cm. With further increasing displacements the forces ceased to be linear and grew at a slower rate than the displacements. There was some indication that the region of transition from elastic "enmeshments" to sliding would extend to amplitudes of the quartz oscillations of a few times 10^{-4} cm.

One interesting result of this investigation is the similarity of both the character and magnitude of the forces of interaction between the surfaces in frictional contact, whether the displacements are tangential or normal. This means that the character of these forces at the stage which precedes sliding, in this case between either a metal or an alloy and quartz, is the same as that of the forces between these materials at the stage which precedes their normal separation. Results similar to those above were also obtained when two metallic surfaces were in " dry frictional " contact; one of these metal plates was placed upon the oscillating quartz crystal with a separating film of glycerine and the other one was placed upon the " dry ", polished, upper surface of the first metal plate. In the presence of a glycerine film the bonds between the quartz and the solid put upon it were apparently more rigid than in the case of contact between the " dry " surfaces, and the " bonded " solid was no longer at rest but took part in the oscillations of the quartz.

§ 7. THE FORCE OF SLIDING FRICTION

It is well known that under certain conditions the sliding motion of elastically restrained bodies does not proceed smoothly but by jerks which exhibit the two characteristics of relaxation oscillations, a fixed amplitude, and a period determined by a relaxation time for the building up of a limiting value. The friction elements appear to " stick together " while the elastic force increases and is balanced by the correspondingly growing force of static friction until the limiting value of static friction is reached, when suddenly a rapid slip occurs (figure 4). In these cases, for small velocities, sliding friction is obviously smaller than the limiting value of static friction, and falls off with increasing velocity. Haykin and his collaborators (1933, 40), and also Blok (1940), have

pointed out that, for jerky motion due to relaxation oscillations to occur, the force-velocity characteristic of sliding of a particular frictional contact must exhibit the known characteristics of any mechanical or electrical device which is capable of performing relaxation oscillations under certain conditions: i.e. a non-linear falling section and a minimum (figure 5). With this type of characteristic the process is either stable or unstable, depending upon the rate of energy supply from a uni-directional source. If this is given a value such that the representative point on the characteristic is to the left of the minimum, conditions are unstable and the system performs relaxation oscillations. If the representative point lies on the rising arc to the right of the minimum, conditions are stable and sliding proceeds without any discontinuity. In the case of a frictional contact stability prevails when $v_0 > v_1$, where v_0 is the uniform velocity at which one of the friction elements is moved forward and v_1 is the sliding velocity for which the frictional force is smallest (figure 5). If $v_0 < v_1$, instability ensues.

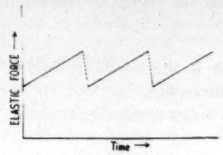


Figure 4.

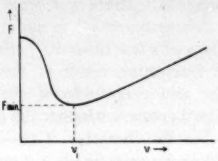


Figure 5.

Figure 4. Force-time diagram when the average velocity of sliding is small, and sliding friction is smaller than the limiting value of static friction. The two friction elements seem to stick together until the elastic force reaches the limiting value of static friction, when rapid slipping occurs.

Figure 5. Friction-velocity characteristic consistent with jerky sliding.

Schnurmann and Warlow-Davies (1942) suggested that the force-velocity characteristic of sliding would exhibit a non-linear falling section and a minimum on account of contact electrification (Ward, 1937; Schnurmann, 1940 b). Whenever two different materials establish contact (Schnurmann, 1941 c) a separation of electrostatic charges takes place at the actual areas of contact which, in general, form a small proportion of the apparent area. In all those cases in which the charges have fixed positions on the surfaces (dielectric materials, and metal surfaces coated with a film of dielectric matter) the charges become separated as sliding proceeds. This gives rise to a tangential electrostatic force (the electrostatic component of the force of sliding friction) which increases as new contacts are made, until the dielectric breakdown field strength is reached. A rapid discharge then follows, and cycles of slow charging and rapid discharging follow one another. If the electrostatic component forms an appreciable proportion of the force of sliding friction, the cycles of slow charging and rapid discharging cause jerky motion, i.e., cycles of slow sliding and rapid slipping. In agreement with this interpretation a discharge current was observed at each jerk, when metallic friction elements, the surfaces of which were coated with a film

of dielectric matter, were connected to a ballistic galvanometer, even when both friction elements were made of steel. In accordance with the view that the galvanometer deflections indicated the discharge current which flowed when the contact potential reached the breakdown value of the boundary layer and ceased when the potential had fallen to a smaller value, for which the resistance of the boundary layer was practically infinite, an appreciably larger galvanometer deflection was observed at the final jerk, because the complete separation of the two friction elements must lead to a *complete* discharge of their surfaces through the galvanometer. Owing to difficulties of arriving at an estimate of the magnitude of the electrostatic component of the force of sliding friction, the order of the normal component of the electrostatic force was computed and compared with the applied normal force. With Helmholtz's assumption that an electrical double layer of about 10^{-8} cm. thickness exists between two dissimilar surfaces which are in contact, and with the further assumption that the difference between the breakdown potential and the truly static value might be one-tenth of a volt, a normal electrostatic force of 4×10^3 kg. weight was estimated and compared with the magnitude of the applied normal force which under similar circumstances amounted to anything between 25×10^3 kg. weight and 30×10^3 kg. weight. It was also possible to show that jerking decreased when the dielectric breakdown strength of the boundary lubricating material was diminished, and that, with the same lubricant, jerking decreased as the average value of the velocity of sliding was increased.

The implication of the electrostatic interpretation of jerky motion, that the cycles of apparent " stick and slip " should be cycles of slow sliding and rapid slipping (slow charging followed by a rapid discharge), could not be tested to the extent of measuring the small amount of slow sliding because, in this transition region from static to sliding friction (see § 6), the internal displacements and forces mask the displacements and forces between the surfaces in contact, when one of the friction elements is elastically restrained and static methods applied.

This interpretation of jerky motion implies that the force-velocity characteristic of sliding for two *naked* metal surfaces in direct contact must have a different shape and cannot contain a falling section and a minimum, because the electrostatic charges are mobile in such surfaces, and the contact potential is proportional to the difference of the work functions and independent of the area of separation of contact. Consequently, sliding between naked *metal* surfaces should be a function of temperature only, and should proceed smoothly even at a small velocity. Thus, smooth relative motion should occur between metals under conditions of severe tearing, under which conditions the plane of displacement is transferred from the true boundary to an initially naked metal-to-metal contact which requires some time to become contaminated. Indeed, in accordance with the electrostatic interpretation of jerky motion due to relaxation oscillations, Schnurmann and Warlow-Davies found that, under conditions of boundary lubrication, Amontons' law was satisfied and the surfaces remained

undamaged even when the tangential force of friction was a large fraction of the normal force, particularly when the jerks were very heavy (figure 6). In contrast to this, when severe tearing was observed, the slow relative motion between the metallic friction elements had proceeded smoothly and did not obey Amontons'

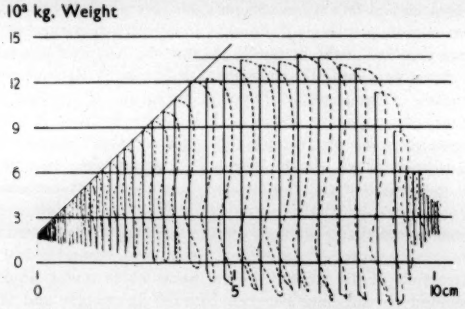


Figure 6. Jerky motion when a steel pin and ring were assembled slowly and without tearing. The frictional force (plotted as ordinate) increased in proportion to the radial force (Amontons' law). The record represents the frictional force as a function of the path travelled by the pin relatively to the ring. The distances between the almost vertical continuous lines indicate the magnitude of slipping. The dotted parts of the record are due to violent vibrations of the recording gear immediately after sudden slipping of the pin into the ring.

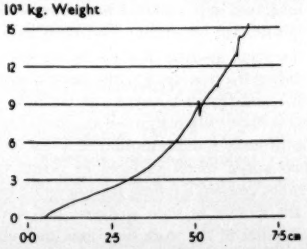


Figure 7. Smooth relative motion and deviation from Amontons' law when relative motion proceeded by severe tearing. The small kinks on the curve were obtained when the propelling operation was suspended. The ordinates represent the frictional force, and the abscissae the magnitude of the path travelled by the pin relatively to the ring.

law. Tearing and deviations from Amontons' law were always observed when either the material used as a lubricant appreciably attacked the metal surfaces, or when these had become discoloured with corrosion before the lubricant was applied. The force of friction then increased at a greater rate than the nominal area of contact (figure 7), and bodily seizure was experienced when the rate of

increase of the force of friction was sufficiently large to lead to a value which exceeded the capacity of the propelling mechanism (Schnurmann, 1940 c). In one of these cases a longitudinal section of the seized elements was prepared and examined under the microscope. This showed that particles had been plucked out from the surfaces and dragged along, but surfaces of separation were clearly visible (figure 8). Also there was no evidence of any appreciable discoloration of the contact area, as would have been expected if the temperature had been high enough for local welding to occur (Schnurmann, 1941 d).

Similar observations of severe tearing between platinum and silver in " dry frictional " contact, for instance, and consequently of the transition from a jerky to a smooth motion, were made by Bowden and Leben (1939), who advanced an entirely different interpretation of sliding which was observed to proceed by a series of jerks. The friction behaviour, in the view of these authors, should depend upon the relative physical properties of the metals, upon the melting point in particular, rather than upon the properties of the boundary layer. Friction should be due to a welding together of the metals at the local places of contact, and the metallic junctions, when broken, should distort the metal " to a considerable depth ". In their detailed discussion of this suggestion, Haykin and his collaborators pointed out, firstly, that " local melting "—if it does take place—would be the result, but not the cause, of jerky motion, and secondly, that " the phenomenon of melting cannot serve as an explanation of the nature of frictional forces for the additional reason that the characteristics of frictional force between wood and metal have in the main the same form as between two dissimilar metals. With a wooden drum on a metal shaft Kaidanovsky and Haykin obtained exactly the same jerky motions as in the case of two dissimilar metals ". Jerky motion was also observed by Papenhuyzen (1938) using rubber on glass, while the jerky motion of a sand column in a glass tube was described as early as 1910 by Phillips, who ascribed it to friction as " the controlling factor ". He conceived the impossibility of obtaining an absolutely uniform *slow* motion using the late Lord Rayleigh's method of loading the sand column with a weight; the downward motion of this weight as the column lowered, due to escape of powder from a nozzle at the end, served to operate a train of wheels. Phillips (1910) made it the object of a lecture experiment to show that " on freeing the nozzle . . . . the sand moves out by slow regular jerks ".

§ 8. THE SURFACE TEMPERATURE OF SLIDING METALS

A conversion of mechanical energy into heat takes place during every process of friction. The loss of mechanical energy can be illustrated by a very elementary experiment. Let a block of solid material, of mass M, rest upon an accurately horizontal table so that its potential energy is independent of the position of the block on the table. If a mass m hanging over a pulley at one end of the table is connected to the block by a taut horizontal piece of string, static friction balances the weight mg if this is smaller than the limiting value of static friction of the block

on the table, and the potential energy of the mass m is mgh, where h is the height of the mass m above some arbitrary reference point. When the block is moved by hand over a distance x towards the pulley, the mass m falls through the height x and its potential energy decreases by mgx to the value $mg\ (h-x)$, while the potential energy of the block remains unaltered. The mechanical energy mgx, which appears to have been lost, has been converted into heat. This frictional heat raises the temperature of the friction elements. The seat of heat generation is the *actual* area of contact, and the temperature which it assumes depends at every instant upon the equilibrium between the rate of heat generation and the rate of heat dissipation, the former being in direct proportion to the speed of sliding and the force of friction, and the latter depending upon the thermal diffusivities of the friction elements, their geometrical dimensions, and the heat dissipation from their surfaces by convection and radiation. The smaller the speed of sliding and the frictional force, and the larger the thermal conductivities, the geometrical dimensions and the emissivities, then, under stationary conditions, the smaller will be the temperature rise at the actual area of contact. Bowden and Ridler (1936) observed that at room temperature a cylinder of 0·1 cm. diameter of an alloy of low thermal conductivity, such as constantan for instance, loaded with 100 g. weight and made to slide with its base on a mild steel surface at a velocity of 100 cm./sec., exhibited a force of sliding friction of 23 g. weight, corresponding to a rate of generation of frictional heat of $5{\cdot}37\times10^{-2}$ cal./sec. From this they calculated that the temperature at the area of contact would be 75° c. above room temperature by assuming that both friction elements would share equally in the dissipation of heat, and that the length of the constantan cylinder was sufficient for its top end to remain sensibly at room temperature. In good agreement with this result Bowden and Ridler measured a temperature rise of about 90° c. at the *actual* area of contact when they used the rubbing contact of the two different alloys as a thermocouple, and determined the electromotive force developed upon sliding when this proceeded at a velocity of 100 cm./sec., and the normal load amounted to 102 g. weight. With gallium (melting point 30·2° c.) sliding on mild steel, a high speed of 720 cm./sec. was required, at a load of 27 g. weight, to raise the temperature at the actual area of contact from 17° c. to 30·2° c., so that melting from the surface of the friction element of lower melting point (in this case gallium) would arise. Having thus shown conclusively that under stationary conditions *large* values of both the force of friction and the speed of sliding are required for achieving an appreciable rise of temperature on the actual area of contact of a solid of high thermal conductivity such as a metal, Bowden and Ridler raised doubt concerning the correctness of their own calculations since, in their view, these had been based on the assumption that the area of contact was that of the end of the cylinder, whereas the actual rubbing area of contact might be " only 1/10, 1/100, or even 1/1000 of the apparent area ". The force of friction, however, takes account of the actual area of contact and, for this and other reasons (Schnurmann, 1939), these criticisms appear to be invalid.

Under the non-stationary conditions of jerky motion, frictional heat is

generated during the short period of rapid slipping which is followed by a comparatively very long period of apparent sticking. A temperature rise at the area of contact of about 1° c. per jerk can be estimated, for instance, in the case of Bowden and Leben's experiment on silver in frictional contact with a steel plate which was propelled at the uniform rate of 6×10^{-3} cm./sec., if each jerk covering a path of $4{\cdot}3 \times 10^{-3}$ cm. was completed within one-thousandth of a second, and if the silver slider can be assumed to be a sphere of 0·3 cm. diameter. For the purpose of this estimate the area of contact between silver and steel was obtained from the value of the Brinell hardness of silver together with the magnitude of the usual load applied in this particular experiment. There is no evidence of " local melting " during jerky motion due to relaxation oscillations between two metals at room temperature. Bowden and Leben regarded the above-mentioned discharge current as a temperature " flash ", but found it difficult to reconcile it with their interpretation of jerky motion. " The apparent temperature rise is small, but two factors prevent us from evaluating the real surface temperature. The temperature flash is so rapid that it cannot be followed by the Einthoven galvanometer. Also the actual points of contact will not all be at the same temperature. Short-circuiting will occur through the colder points and the e.m.f. recorded will only be a mean value ". It appears that there is certainly no direct experimental justification for the suggestion of *localized* high temperatures on a metal surface (also see Morgan, Muskat and Reed, 1941).

§ 9. THE COEFFICIENT OF FRICTION

Relative motion between two solids in frictional contact may proceed either by sliding or by tearing. Under certain boundary conditions a solid may adhere to an inclined plane so that the limiting value of the force of static friction exceeds the normal load; no sliding will then take place even at an angle of inclination of 90°, and the coefficient of static friction appears to be *infinite*, although a *finite* tangential force would suffice to produce relative motion because the shear strengths of the materials are finite.

A few simple experiments may illustrate, firstly, that strong adhesion between two solids in air can be ascribed to the tenacity of a monomolecular boundary layer and is encountered under conditions under which the formation of welded junctions is not feasible, and secondly, that, under conditions of either " dry friction " or boundary lubrication, the value of the coefficient of friction primarily depends upon the thickness and nature of the boundary film, so that the same solid materials may exhibit coefficients of static friction covering the whole possible range from almost zero to infinity. Various procedures may be chosen for reducing the thickness of the boundary layer; they all lead to similar results. The wringing of end gauges is well known. Similarly, Lord Rayleigh (1936) has discussed the method of getting glass or silica plates into optical contact and has measured the pull required to separate two such glasses; the highest value found was 46 kg. weight/cm^2. He described in detail the extrusion of the

major part of the boundary film when two glass plates of different size, both with chamfered edges, were worked together with circular sweeping strokes. Most of the benzene with which the larger surface had been flooded before the other was brought down upon it was extruded in this way, and the friction rapidly became greater. When further movement became difficult, the glasses were set aside for a few hours for the residual benzene to evaporate. At the end of that time, the glass plates were found to be in contact over part, at any rate, of the area and to be adhering so firmly that it was difficult to slide them one over the other. Similar results were obtained when a small metal cylinder was cast on to a mechanically clean fused glass plate, and the greater part of the boundary film was "boiled off" the area of contact on the glass plate instead of being removed by extrusion (Schnurmann, 1942). The solidified metal cylinders adhered so as to carry a nominal normal stress of about 15 kg. weight/cm. in the case of Wood's metal and of about 25 kg. weight/cm.^{2} in the case of lead. The strong adhesion also became apparent from the convex shape which the surface of the solidified metal cylinder assumed upon being separated from the glass plate. It was evident that the hot metal, poured at a temperature close to its melting point, drove off, from the area of contact of the glass plate, most of the moisture which was present there at room temperature. The areas of contact of metals other than Wood's metal had a mottled appearance (figure 9) and contained numerous blowholes; these blowholes were less numerous and appreciably smaller when the glass plates had been heated to 200° c. and more, in a furnace with air circulation, before cylinders of lead were cast on to them while they remained in the furnace (figure 10). Evidently the adhesion between the metal and the glass must arise at the islands of metal between the blowholes, although it cannot be attributed to the formation of bridges of *direct* contact between metal and glass, since the islands show etched structure (figure 11). It was thought, therefore, that the driving off of a large fraction of the moisture from the area of contact on the glass plate, and the consumption of part of the residual moisture for etching the islands would leave a sufficiently thin film between the glass and the islands of metal to act as an adhesive; it was found that, owing to the larger thermal conductivities of metals, cylinders cast on to metal plates would not adhere unless these plates had been sufficiently preheated for most of the moisture to be driven off before the molten metal was poured.

Further evidence of the adhesive action of thinned-out boundary layers was obtained with porous materials which exerted a blotting action. Of particular interest in this connection is solid H_2O in its two modifications of ice and snow. The reason why snow adheres to an appreciably inclined base which is at a temperature above the freezing point of water, whereas a block of ice slides on the same base at a very small angle of inclination, is that the porous snow blots the major proportion of water formed by melting from the area of contact, leaving a sufficiently thin layer of melt for adhesion between the snow and the solid, whereas melting from the area of contact of a block of ice provides thick-film lubrication

Figure 8. Photomicrograph (magnification about 400 times) of the longitudinal section of two seized friction elements of steel, showing surfaces of separation.

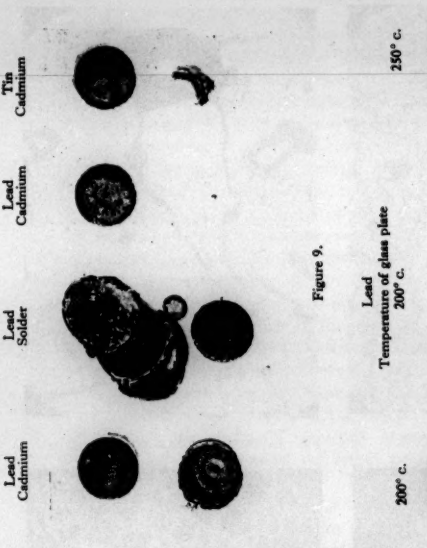


Figure 9.

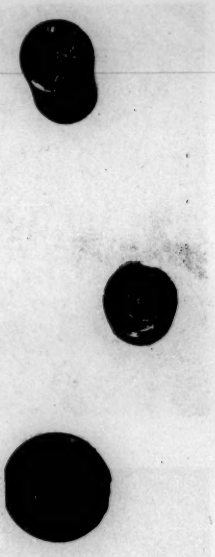


Figure 10.

Glass plate baked out at 500° c. Lead cylinder cast when glass plate has cooled down to 230° c.

Figure 10. Apparent areas of contact of cylinders of lead cast on to glass plates which were at various temperatures, in air.

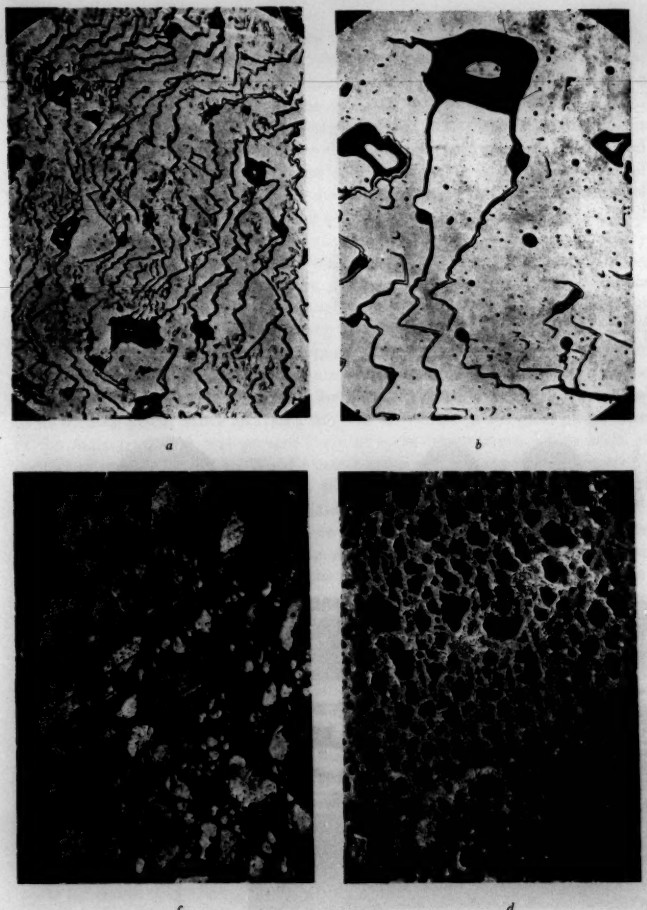


Figure 11. *a* and *b*. Etched structure (magnification about 83 times) of an " island of contact " of a cylinder of lead adhering to a glass plate which was at room temperature and at 230° C. respectively when the molten metal was cast on to it. *c* and *d*. A part (magnification about 83 times) of the material remaining on a glass plate from which cylinders of cadmium and of eutectic Cd–Zn alloy respectively have cracked off.

(Schnurmann, 1940 d). It is well known that an appreciable load of snow will adhere to the inclined roof of a house as long as the rate of melting of the snow at the area of contact does not exceed the rate at which water can be blotted by the bulk of the snow. The blotted water is then transferred from the interface between the snow and the roof, flows through the snow capillaries and forms icicles if the temperature of the air is below the freezing point of water, so that melting occurs at the area of contact only. The coefficient of static friction in this case depends both upon the rate of melting at the area of contact and upon the rate of blotting of the molten material, and assumes, for instance, the value 0·7 if a lump of snow has been compressed by hand before being put upon an inclined steel plate which is at room temperature, while the coefficient of loose snow would appear to be infinite.

In a different experiment, in which mercury was made to slide over a glass surface, direct evidence was obtained that the thickness of a film of condensed water vapour was monomolecular when it exerted appreciable adhesive properties.

The coefficient of static friction between mechanically clean surfaces in air appears to be very small when the boundary film of adsorbed and condensed matter is sufficiently thick to allow of easy shear. In the case of an inclined-plane experiment this coefficient generally appears to be small if the ratio between load and shear strength for either component of the combination solid/boundary film/solid is very large. Sliding will begin by tearing one of the solid surfaces if this is the weakest part of the joint.

Shaw and Leavey observed that the coefficient of static friction measured by the tilt method was much higher, especially for light loads, when the friction elements were in a chamber which had been evacuated to about 10^{-2} mm. of mercury than when they were treated in open air, and both Holm and Kirschstein (1936, 39) and Bowden and Hughes claimed that the coefficient of friction of outgassed metals " might be twenty times greater than that observed for the same metals in air ". It appears, however, that the results obtained in air may be either smaller or larger than the coefficient of solid friction (between naked surfaces), depending upon the boundary conditions. The experiments on solid friction to which reference was made in § 5 revealed a coefficient of static friction of the order of unity at room temperature when the surfaces had been cleaned by volatilization in a high vacuum, while out of contact, at a temperature at which the whole friction chamber was baked out. In cases in which the latter condition could not be satisfied, for instance with friction elements of steel, the whole friction chamber was baked out at a temperature of 450° c. and a pressure of 10^{-6} mm. Hg. Thus, a coefficient of static friction of 0·9 was obtained for steel at room temperature, and a gradual decrease to between $0{\cdot}5_5$ and 0·6 was observed. Fogg and Hunwicks (1940), who resorted to the drastic procedure of removing completely, in air, the contaminated surfaces of friction elements of steel by abrasion, observed coefficients of static friction between 0·56 and 0·60 when the measurements were made as quickly as possible after the abrasion treatment. Coefficients of

static friction of the order of unity were also found with freshly split mica surfaces (Derjaguin and Lazarew, 1934), and with metals when the surfaces suffered tearing at a pressure of about 10^{-1} mm. Hg. in the friction chamber (Ernst and Merchant, 1940).

A transition from a lower value of the coefficient of friction of about 0·6 to about 1·3 in the course of relative motion between platinum and silver in air at atmospheric pressure and temperature was observed by Bowden, Leben and Tabor (1939) when a blob of silver became attached to the platinum slider. Interesting results with regard to the force-velocity characteristic of sliding were published recently by Beeck, Givens and Smith (1940). They used the four-ball apparatus, in which a single ball rotates under variable loads on a support formed by three similar balls clamped together in an oil cup. At low sliding velocities of 0 to 1 cm./sec. and high pressures of a few thousand kg. weight/cm^2 or more the coefficient of friction-velocity diagrams, for various boundary lubricants, were determined by means of deceleration experiments. With a white oil the non-linear falling section to which reference has been made in §7 (figure 5) was found, whereas the characteristic of ricinoleic acid showed hardly any dependence upon the velocity of sliding. The majority of the boundary lubricants showed, however, a rapid decrease of the coefficient of friction at the initial higher velocities. With decreasing velocity, these curves bend over sharply at certain critical velocities, and the coefficients of friction become constant, as can be expected if the type of the force-velocity characteristic changes upon a transition from sliding to tearing. The authors offer an alternative interpretation of their results, and assume the wedging of oil under the surface (oil drag), leading to a " quasi-hydrodynamic " type of lubrication in the presence of long-chain polar compounds. They found that changing the viscosity of the white oil by blending it with a non-polar material had little effect. Such blends, containing a non-polar material such as polyisobutylene and with 1 per cent of oleic acid, whose viscosities differed by a factor of 100, showed only about 10 per cent change in the critical velocity and in the coefficient of pure boundary lubrication. These authors concluded, therefore, that differences in bulk properties of the lubricant would have a negligible effect upon the " wedging ".

REFERENCES

Amontons, G., 1699. *Mém. de l'Acad. Roy. des Sciences*, p. 206.
Beeck, O., Givens, J. W. and Smith, A. E., 1940. *Proc. Roy. Soc.* A, **177**, 90.
Bikerman, J. J., 1941. *Phil. Mag.* **32**, 67.
Bikerman, J. J. and Rideal, E. K., 1939. *Phil. Mag.* **27**, 687.
Blok, H., 1940. *J. Soc. Aut. Eng.* **46**, 54.
Borelius, G., 1935. *Handbuch der Metallphysik*, **1**, 369.
Bowden, F. P. and Hughes, T. P., 1937. *Proc. Roy. Soc.* A, **160**, 575.
Bowden, F. P. and Hughes, T. P., 1939 a. *Proc. Roy. Soc.* A, **172**, 280.
Bowden, F. P. and Hughes, T. P., 1939 b. *Proc. Roy. Soc.* A, **172**, 263.
Bowden, F. P. and Leben, L., 1939. *Proc. Roy. Soc.* A, **169**, 371.
Bowden, F. P., Leben, L. and Tabor, D., 1939. *Engineer, Lond.*, **168**, 214.

BOWDEN, F. P. and RIDLER, K. E. W., 1936. *Proc. Roy. Soc.* A, **154**, 640.
BOWDEN, F. P. and TABOR, D., 1939. *Proc. Roy. Soc.* A, **169**, 392.
COULOMB, A., 1785. *Mém. de l'Acad. Roy. des Sciences*, p. 161.
DERJAGUIN, B., 1934. *Z. Phys.* **88**, 676.
DERJAGUIN, B. and LAZAREW, W., 1934. *Koll. Z.* **69**, 11.
ERNST, H. and MERCHANT, M. E., 1940. *Proc. of the Spec. Summer Conferences on Friction and Surface Finish* (*Mass. Inst. Technology*), p. 76.
ESTERMANN, I. and STERN, O., 1930. *Z. Phys.* **61**, 95.
FOGG, A. and HUNWICKS, S. A., 1940. *J. Inst. Petroleum*, **26**, 1.
FRENKEL, J., 1926. *Z. Phys.* **37**, 572.
HARDY, Sir W. B. and DOUBLEDAY, I., 1922. *Proc. Roy. Soc.* A, **100**, 550.
HAYKIN, S., LISSOVSKY, L. and SOLOMONOVICH, A., 1939. *J. Phys.* **1**, 455.
HAYKIN, S., LISSOVSKY, L. and SOLOMONOVICH, A., 1940. *J. Phys.* **2**, 253.
DE LA HIRE, P., 1699. *Histoire de l'Académie*, p. 104.
HOLM, R. and KIRSCHSTEIN, B., 1936. *Wiss. Veröff. aus den Siemens-Werken*, **15**, 122.
HOLM, R. and KIRSCHSTEIN, B., 1939. *Wiss. Veröff. aus den Siemens-Werken*, **18**, 73.
JACOB, C., 1912. *Ann. Phys., Lpz.*, **38**, 126.
KAIDANOVSKY, N. L. and HAYKIN, S., 1933. *J. Techn. Phys.* **3**, 91.
MERCHANT, M. E., 1940. *J. Appl. Phys.* **11**, 230.
MORGAN, F., MUSKAT, M. and REED, D. W., 1941. *J. Appl. Phys.* **12**, 743.
PAPENHUYZEN, P. J., 1938. *De Ingenieur*, **52**, 75.
PHILLIPS, C. E. S., 1910. *Proc. Roy. Inst.* **19**, 742.
PRANDTL, L., 1928. *Z. angew. Math. Mech.* **8**, 85.
RANKIN, J. S., 1926. *Phil. Mag.* **2**, 806.
RANKIN, J. S., 1928-29. *Trans. Sci. Soc. Roy. Techn. Coll., Glasgow*, **15**, 1.
RAYLEIGH, Lord, 1936. *Proc. Roy. Soc.* A, **156**, 326.
SCHNURMANN, R., 1938. *Mech. Engng.* **60**, 259.
SCHNURMANN, R., 1939. *Engineer, Lond.*, **168**, 278.
SCHNURMANN, R., 1940 a. *J. Appl. Phys.* **11**, 624.
SCHNURMANN, R., 1940 b. *Proc. Phys. Soc.* **52**, 179.
SCHNURMANN, R., 1940 c. *Engineering*, **149**, 567; **150**, 236.
SCHNURMANN, R., 1940 d. *Nature, Lond.*, **145**, 553.
SCHNURMANN, R., 1940 e. *Mech. Engng.* **62**, 408.
SCHNURMANN, R., 1941 a. *Proc. Phys. Soc.* **53**, 538.
SCHNURMANN, R., 1941 b. *Proc. Phys. Soc.* **53**, 662.
SCHNURMANN, R., 1941 c. *Proc. Phys. Soc.* **53**, 547.
SCHNURMANN, R., 1941 d. *J. Iron Steel Inst.* **143**, 415 P.
SCHNURMANN, R., 1942. *J. Appl. Phys* **13**. In press.
SCHNURMANN, R. and WARLOW-DAVIES, E., 1942. *Proc. Phys. Soc.* **54**, 14.
SHAW, P. E. and LEAVEY, E. W. L., 1930. *Phil. Mag.* **10**, 809.
STEVENS, J. S., 1899. *Phys. Rev.* **8**, 49.
TOMLINSON, G. A., 1929. *Phil. Mag.* **7**, 905.
DA VINCI, LEONARDO, *c.* 1500. *Il Codice Atlantico* (ed. by Accademia dei Lincei, Milan, 1900–3).
WARD, W. H., 1937. *Rep. Prog. Phys.* **4**, 247.

THE GENERAL PHYSICAL CONSTANTS
AS OF AUGUST 1941 WITH DETAILS ON THE VELOCITY OF LIGHT ONLY

By RAYMOND T. BIRGE,
University of California, Berkeley, California

INTRODUCTION

THIS paper is being written on request—and at *this time* on request. In spite of the delusive word " Constants " appearing in the title, it is the continual variation in the values of these quantities that furnishes most of the interest in the subject. It would indeed be disheartening to any real scientist to feel that an accepted value of any physical constant would never again be changed. The most characteristic feature of science—in general and in particular—is its never-ending change. But what one must not overlook is the fact that it is the *probable* value of any given constant that changes, not its *actual* value. A belief in any significant variability of the constants of nature is fatal to the spirit of science, as science is now understood.

It is, of course, a fact that the experimentally determined value of every " constant " varies in general with each determination, and just such variations often furnish invaluable clues to errors of experiment and of theory. It has always seemed to me that the chief value of the critical work that has been done on the general constants lies just in its potential exposure of such experimental or theoretical errors. In fact, possibly as a result of such critical work, the situation in respect to the general constants is now vastly improved over that obtaining in 1929, when I wrote my first general paper (Birge, 1929; hereinafter referred to as G.C. 1929) in the field; and for the first time since 1929 it seems possible to present with some confidence a new set of values.

Another most desirable object of a critical study of the general physical constants is to furnish a *consistent* set of values. This object, no matter how desirable, is really impossible of attainment, and the longer one works in the field, the more one is impressed with this rather discouraging fact. The general physical constants are woven together in a net of innumerable strands. It is scarcely an exaggeration to say that any change in the adopted value of a given constant affects, *to a greater or less degree*, the apparent best value of every other constant. Fortunately this degree of dependence varies very greatly from constant to constant, and in G.C. 1929 I attempted to reduce the difficulty to a minimum by starting with those least dependent on others, and working finally to those dependent on others.

Because of the general interdependence of all constants, it is never possible to work up a strictly consistent set of values that at the same time represents the best experimental evidence. The work on any such paper as this necessarily extends over many months, if not years, and during the work new experimental results are constantly appearing—not to mention the *old* results that one, in some inexcusable way, has overlooked. Any one such new result, if accorded proper recognition, is capable of spoiling half the calculations of an entire paper. It may not be inappropriate to mention also the human element involved. One should, in principle, undertake such work very methodically, beginning, as just stated, with the constant least dependent on others, and working systematically through the remaining constants, in a carefully predetermined order. Actually, however, one's time is strictly limited, especially in these days of international stress, and when an exciting new result appears, one is apt immediately to test its effect on the most probable value of the constant concerned, without stopping to investigate whether any of the auxiliary constants have also recently changed in value. Such a calculation frequently develops into an extensive investigation, which one would prefer not to repeat too often.

A particularly unfortunate example of such a situation may be mentioned. The values of the electronic charge e and the specific electronic charge e/m appear more often in the numerous derived constants than possibly any other of the fundamental constants. Now the value of e depends on the value of the Faraday F, when e is calculated in the present most reliable way. Just one half of the present twelve most reliable values of e/m also involve F. But the value of F depends upon the adopted atomic weights of iodine and silver, and the most accurate determinations of I depend in turn on the value of Ag. Since 1925 the internationally accepted value of Ag has been 107·880, and there appears to be nothing in any recent report of the International Committee on Atomic Weights to indicate that any change is contemplated or needed in the adopted value. Hence, in working up the present list of constants, I merely accepted the value $107{\cdot}880 \pm 0{\cdot}001$, which I had adopted in G.C. 1929. Then, after my calculations were completed, I happened to read something in my notes that aroused my suspicions about the value of Ag. As the result of a subsequent hasty and still incomplete investigation, it appears to me that the best value is far more likely to be 107·878 than 107·880. To get out of this difficulty with a minimum of recalculation, I have increased the probable error and now *adopt* $\mathrm{Ag} = 107{\cdot}880 \pm 0{\cdot}002$, thus covering the possibility of 107·878. Since the atomic weight of silver is the basis for nearly all atomic weights, it is not to be changed without due consideration.

The moral of the incident is that one should take nothing for granted. Unfortunately no one can be an expert in every field, and I often feel that I should apologize for expressing critical opinions in many fields concerning which I know very little. It is true, however, that the experts in one field often use values of auxiliary constants definitely different from those used by experts in another

field. In fact, all too often the various experts in a given field are using simultaneously a bewildering variety of values. Strangely enough, such variations are often marked off by strict geographical boundaries, a state of affairs that has shown vast improvement in the past century, but is still open to a lot more improvement. Hence, in order to arrive at any *consistent* set of values, it is necessary to go far " behind the scenes " of any published result, and just because of the practical impossibility of doing this either thoroughly or simultaneously for all the constants, no published list of constants can ever be satisfactory to the author, regardless of how the reader may feel.

With this very personal, but I trust pertinent, explanation of the problems involved in critical work on physical constants, I proceed to the details. Within the limits placed on the present article, it is possible to set forth only a very small fraction of the relevant material. I hope later to publish elsewhere a far more extended account. Hence I have chosen just one of the general constants for a detailed discussion—namely, the velocity of light. This constant furnishes a good illustration of the sort of problems that arise in the critical evaluation of any constant, and because of the much publicized suggestion that the velocity of light may be changing in value, I trust that the discussion will be not without interest.

On the other hand, I have now practically finished a re-examination of all the general constants, and the newly recalculated and adopted values are given in tables at the end of this paper, together with the adopted values of all auxiliary constants used in the calculations, and the values of a few derived constants. A very brief statement is made concerning each of the general constants, and the lettering of sections is the same as that used in G.C. 1929. Incidentally, the introduction to that paper should be considered also as an additional introduction to the present paper.

SECTION A

THE VELOCITY OF LIGHT IN VACUUM (c)

The velocity of light is possibly the most fundamental of all constants. Certainly more time and money have been devoted to its measurement than in the case of any other general constant. But in spite of its long history, beginning with Römer in 1675, no measurement comparable in accuracy with recent work was made previous to 1900. Science moves forward at an ever increasing pace, and three-quarters of the papers to be cited in this article have been published since my previous account in 1929 ! But de Bray (1927–34) appears to take quite seriously certain older measurements, and as a result he obtains the truly alarming deduction that the velocity of light is decreasing at the rate of some four km./sec. per year. At that rate the velocity would go to zero in a mere 75,000 years. Now it is quite possible that the length of the day is increasing very slowly—let us say 0·01 sec. per century. This change would cause the apparent velocity of light to increase one part in 10^7 per century, whereas the de Bray equation

corresponds to a *decrease* of one part in 750 per century. Since this question of the variability of c has attracted much attention, among both scientists and laymen, it is advisable to consider the present evidence rather critically.

In G.C. 1929 I merely accepted Michelson's 1927 result of $c = 299796$ km./sec. (Michelson, 1927). This result was believed, at the time, to have a probable error of only 4 km./sec. But Michelson, Pease and Pearson (1935; to be denoted by M.P.P.) found irregular and completely mysterious variations in the measured value of c occurring daily, and even hourly, as well as over longer periods of time. A subsequent examination of Michelson's original notebooks, I am told, has shown that similar variations were present in his work, although he had failed to notice them at the time.

Michelson greatly overestimated his accuracy in many respects. Thus the distance from Mt. Wilson to Mt. San Antonio (" Baldy "), some 35 km., was believed to have been measured correctly to about one part in 10^6 (or 35 mm.). On the other hand, M.P.P. found that their carefully measured mile-long base line showed for two years a steady increase in length of 6·5 mm. per year (or 4 parts in 10^6 per year), and then a sudden *decrease* of 8 mm. after a mild earthquake in the vicinity! As the late D. C. Miller has noted, between the time that Michelson's 35-km. base line was measured and his actual measurements on c were made, there occurred the disastrous Santa Barbara earthquake, and his actual base line may well have suddenly changed by several feet! Furthermore, Michelson's measurements were made under a wide variety of temperatures and barometric pressures, but Michelson used only a single correction to vacuum for his measured velocity in air. For these, and other reasons, I am now increasing the probable error of Michelson's 1927 work to 15 km./sec. This new probable error may be slightly too large, and this point will be mentioned again later.

Michelson's published result of 299796 km./sec. is also in need of revision. To obtain it he used 67 km./sec. for the correction to vacuum from air at an assumed average temperature of 20° c. and at an average barometric pressure of 625 mm. Hg. It is easy to show that his correction results from the use of the *wave* index of refraction, instead of the correct *group* index of refraction. This is one of the most inexplicable errors that I have ever come across in the literature. It is well known that Michelson, about 1883, obtained 1·330 ($\pm$?) for the index of refraction of water, and $1{\cdot}76 \pm 0{\cdot}02$ for the index of CS_2, by a direct comparison of the velocities in air and in the two media mentioned.

Now the *wave* index of refraction of CS_2 is 1·624 for the " D " line, and 1·637 for the " E " line. To explain his high observed value for CS_2, Michelson employed the concept of *group* velocity (and the corresponding *group* index of refraction), as discussed in connection with the velocity of light by Rayleigh (1881 a, 81 b), although the first correct formulation of group velocity should be credited to W. R. Hamilton (*c.* 1835), according to a review in *Nature* (Ramsey, 1940). For CS_2 the group index μ_g ($=\mu - \lambda_0 d\mu/d\lambda_0$, where μ is the wave index and λ_0 the wave-length in vacuum) is 1·722 and 1·767 for the " D " and " E "

lines, respectively, in good agreement with Michelson's observed result. This is, however, the *only* instance in which Michelson ever used the concept of group velocity! He never applied it to his work on the absolute value of c, and neither has anyone else working in the field, until I called Dr. Anderson's attention to the matter (see Anderson, 1941) several years ago. Michelson did not even apply group velocity to his observed index for water. If he had, it would have spoiled the apparent agreement, for the group index is 1·352 and 1·359 (for the D and E lines), whereas the wave index is 1·334 and 1·336, as compared to his observed 1·330.*

The correction of the velocity of light to vacuum from air, for an effective wave-length of λ5700, is 82·97 km./sec., at 0° c., 76 cm. Hg. pressure, with the use of the wave index of refraction, and 85·64 km./sec. with the use of the correct group index. At Michelson's pressure and temperature the differential correction is slightly over 2 km./sec. Hence Michelson's actual 1927 result is 299798 km./sec., in place of his 299796. I first noted these facts regarding group velocity shortly after I had written my letter to *Nature* on the value of c (Birge, 1934 b). Since no such correction was needed in the work of M.P.P., and since their work had then shown that Michelson's 1927 work was far less reliable than had been thought, there seemed to be no proper excuse for publicizing the matter. But the group versus wave-index correction becomes more serious in the case of recent Kerr cell work on c, as will be discussed presently.

We come now to the work of M.P.P. This is one of two really extensive recent investigations on the value of c, carried out under controlled and accurately known conditions. The second investigation is that of Anderson in 1940, in which a Kerr cell was used (Anderson, 1941). Although Anderson's work will be mentioned in connection with other Kerr cell work, it is convenient to consider at this point certain aspects of the work of both M.P.P. and Anderson.

The measurements of M.P.P. were made in a mile long evacuated tube, under definitely known and very steady conditions of temperature and pressure. A total of 2885 values of c were obtained from observations made on about 165 different days (actually *nights*), extending over 24 months' time. Anderson obtained 2895 values of c from observations made on 22 different days, extending over 18 months' time. The average value of all the M.P.P. observations is 299774 km./sec., and the average of Anderson's is 299776. The mean epoch of the M.P.P. work is 1932·5, that of Anderson is 1940·0.

It is not possible to make an entirely satisfactory comparison of the consistency

* Similarly Michelson found that "orange-red" light travels 1 to 2 per cent faster than "greenish-blue" light in CS_2, a result that he considered, and that has since been regularly quoted as, in good agreement with theory. Now if we take 6000 A. and 5000 A. as reasonable values for the effective wave-lengths of the light used, the expected difference in velocity is 1·4 per cent on the basis of the wave index, but 3·95 per cent on the basis of the correct group index. Gutton (1911) has shown conclusively that the *dispersion* of the velocity of light in CS_2 does agree with the group-index equation, and *not* with that for the wave index. Gutton obtained no absolute values of the velocity.

of the two sets of measurements, partly because the form of the published material is not the same in the two cases. M.P.P. publish data from which one can plot a distribution curve of the 2885 residuals. They also show such a plot. Most fortunately the plot is quite symmetrical, so that the average value of $c = 299774$ km./sec. can be accepted with some confidence. On the other hand, the distribution of the residuals deviates appreciably from a normal error curve. But I find that a very good fit can be obtained if one takes the sum of two such curves, one with a standard deviation of 5 km./sec. (for a single observation) and the other of 15 km./sec.* Now 5 km./sec. is roughly the deviation to be expected for a single observation. Hence we can assume that in the case of one half the observations, the expected experimental conditions were attained, whereas in the case of the other half, a multitude of errors of one kind or another led to a standard deviation three times as large. Of course many other similar assumptions would doubtless give an equally satisfactory fit. The point is that if a large assortment of observations are *not* of equal reliability, their residuals cannot be expected to follow a normal error curve, and I am more and more convinced that the deviations from such a curve found so commonly in large groups of physical measurements are due usually just to this cause.†

M.P.P. tabulate their observations in 233 series. The average deviation *within* each series from the mean of that series is about 10·5 km./sec. A plot of their observations against time shows unmistakable evidence of irregular deviations from the general average of 299774 km./sec. Deviations of 10 km./sec., over a period of a week or more, are common, and deviations up to 30 km./sec. occasionally occur. The authors were unable to locate the precise origin of these variations. They are probably due in part to instrumental difficulties and in part to the fact that the earth must be considered as a dynamic rather than as a static body.

Anderson gives only the 22 average values for his 22 days of observing. The average deviation *within* each of the 22 series is 9 km./sec., comparing well with the M.P.P. figure of 10·5 km./sec. The average deviation of the 22 daily averages from the final average is 14 km./sec. A corresponding result for the M.P.P. data could be derived, by sufficient calculation, but thus far I have computed only the 13·37 km./sec. standard deviation‡ of the entire 2885 observations, as already stated. It would appear, however, that the distribution of residuals for the two investigations has essentially the same average spread, although I have no information on the *shape* of Anderson's distribution curve.

Anderson himself believes that the irregular deviations with time occurring in his work are due chiefly to a certain instrumental difficulty, which he discusses

* The best single normal error curve that can be passed through the data has a standard deviation of 13·37 km./sec.

† But with carefully controlled experimental conditions, the writer (Birge, 1932 a) obtained an entirely satisfactory normal distribution of residuals.

‡ The average deviation is the arithmetic average, whereas the standard deviation—which gives a more reliable estimate of uncertainty—is the root-mean-square average.

in detail. Just because of the known existence of this difficulty, I shall assign a larger probable error to his final result than to that of M.P.P. With a standard deviation of 13·37 km./sec. for 2885 observations, the probable error of the average is, by formula, only 0·17 km./sec. Obviously, as a result of the systematic and instrumental errors that were evidently present, one must assume for the *effective* probable error of the average value of c a quantity far greater than this. I think, however, that 4 km./sec. is a reasonable estimate in the case of the M.P.P. work, and with this figure as a basis I take 6 km./sec. as the probable error of Anderson's 1940 work.*

Because of the low pressure of the air (0·5 to 5·5 mm. Hg) in the M.P.P. vacuum tube, the total correction from air to absolute vacuum is only 1·5 parts in 10^8, on the average, and the distinction between wave and group velocity is obviously negligible. Anderson (1941) has, however, suggested that the M.P.P. result may need correction for group velocity, due to the fact that the light enters and leaves the vacuum tube through a glass window (0·75 inch thick). Now the change in their total optical path of ten miles, resulting from the use of group velocity in place of wave velocity through 1·5 inches of glass, is 0·073 inch, or only one part in nine million. Hence the final M.P.P. result is $c = 299774 \pm 4$ km./sec. Anderson used the group-velocity equation in getting his published average of 299776, and hence I adopt 299776 ± 6 km./sec.

We now consider as a group four investigations on the velocity of light, made by means of a Kerr cell, which replaces the toothed wheel in Fizeau's classic arrangement. The base line in each case was about 40 metres. One of the four investigations is Anderson's 1940 work, just discussed. In each of the four investigations a quite distinctive experimental arrangement was employed, but to present an adequate discussion of these different arrangements would require many pages. Due to limitations of space I am omitting, in general, all reference to experimental details. On the other hand, it is essential to consider the corrections to the various published values, required by the substitution of group velocity for wave velocity.

The first precision Kerr cell work is that by Mittelstaedt (Karolus and Mittelstaedt, 1928; Mittelstaedt, 1929 a, 29 b). His light path of 250 metres included a portion of the nitrobenzene in his Kerr cells, as well as glass lenses and, of course, air (at 75 cm. Hg average pressure). Mittelstaedt states that in order to get the true optical path length, the geometric length should be increased by 68 mm. for the air, 21 mm. for the glass and 12 mm. for the nitrobenzene, a

* There are two reasons for assigning a probable error. The first is to give an actual *estimate* of the size of the region within which the odds are even that the true value lies. The other, and more important reason for our present purpose, is to get a basis for the *relative weights* to be assigned the various observed values, when calculating a final weighted average. For this second purpose only the *relative* probable errors are of concern. In the case of the velocity of light, I take ± 4 km./sec. for the M.P.P. work as my basic probable error, and I have then attempted to assign such probable errors to other results as seem reasonably to express the relative reliability of the various investigations. This general principle holds throughout the present paper.

total of 101 mm. These results, as usual, correspond to the wave index, for which I find that the exact correction is 68·36 mm. for 250 metres of air, at 75 cm. Hg pressure and for λ 5500 A., the effective wave-length used by Anderson and by Mittelstaedt. The needed dispersion constants are listed by Anderson (1941), and I find for the proper corrections, using the *group* index of refraction, 70·75 mm. for air, 22·82 mm. for glass, and 13·96 mm. for nitrobenzene, a total of 107·53 mm.* The total correction to c is, therefore, $(101/250000) \times 299776 = 121{\cdot}11$ km./sec., according to Mittelstaedt, but is correctly $(107{\cdot}5/250000) \times 299776 = 128{\cdot}94$ km./sec., an increase of 7·83 km./sec. Thus Mittelstaedt's true result is 299786 km./sec., in place of the figure 299778 that he publishes. Mittelstaedt gives 20 km./sec. for his "limit of error". I still believe, sa I did in 1934 (Birge, 1934 b), that 10 km./sec. is a reasonable value for his probable error.

The second Kerr cell result to be discussed is that of Hüttel (1940). In his experimental set-up, which represents a great improvement over that used by Mittelstaedt, there is only air in the light path. The necessary correction to his published result is therefore presumably about 3 km./sec. (2·9 km./sec. for λ5500 A., or 2·67 km./sec. for λ 5700 A.). His published result is 299768, and I correct this to 299771 km./sec. This result is the average of 135 observations, listed in 8 groups that show a maximum variation of 35 km./sec. Hüttel adopts 10 km./sec. for his final probable error, and since his work is certainly as reliable as that of Mittelstaedt, I accept his estimate.

The Kerr cell work of Anderson (1937, 41) has been carried on over more than three years. In his first experimental arrangement (1937), a certain amount of glass (lenses) was included in the light path. In his second paper (1941) he gives the group-velocity correction, to which I had meanwhile called his attention. The correction is 7 km./sec. for his first investigation, making his result 299771 km./sec., in place of 299764. To obtain this average value Anderson made 651 measurements of c, and his plot of their distribution shows approximately the same "spread" as in the case of the M.P.P. work. The observations were taken on 12 different days, distributed over 6 months of time. The average deviation within each of the 12 daily averages is 12 km./sec., and these 12 averages have an average deviation of 8 km./sec. from the final average. Because of the small number of observations in Anderson's first investigation, as compared to that in his second, I assign a probable error of 10 km./sec. to his first average, just as I have done in the case of Mittelstaedt and of Hüttel. Anderson's first result is, then, 299771 ± 10 km./sec., and his second, as already discussed in detail, is 299776 ± 6 km./sec. In the second investigation, all glass lenses were eliminated from the light path, and the electrical arrangements were quite different.

There are two indirect methods for measuring c, although neither method is more indirect than *any* method for measuring h/e, or any other atomic constant.†

* Anderson in some way calculates only 13·4 mm. for the nitrobenzene.

† de Bray appears to ignore such indirect methods in his discussions of the value of c.

The first method is the measurement of the velocity of electromagnetic waves, guided by parallel wires. This method assumes merely that visible light waves and electromagnetic waves have the same velocity in vacuum. The only real precision work with this method is that by Mercier (1923 a–24). In the case of parallel wires large corrections are required to reduce the observed velocity to the velocity of free waves in air. An elementary account of the underlying theory is given in a recent text by Suydam (1940). Mercier's final average—and the only final result that he gives—is 299700 km./sec. I quoted this figure in G.C. 1929 as the resulting velocity in vacuum. Actually, however, it is the velocity in ***air***, and when multiplied by the square root of the dielectric constant of air ($\epsilon = 1{\cdot}000545$ at 20° c. and one atmosphere) one gets 299782 km./ sec. as Mercier's true value for vacuum.

Although the calculated corrections just mentioned vary from 130 to 366 km./sec., for Mercier's five listed results, the corrected results themselves show a maximum deviation from the average of only 30 km./sec., thus indicating the essential correctness of the theory. Mercier takes the 30 km./sec. deviation as his " limit of error ", but because he has thus made no allowance for systematic errors of either theory or experiment, I will assume 30 km./sec. as his ***probable error***, just as I did in 1934 (Birge, 1934 b).

The second indirect value of c is that given by the ratio of an electric charge measured in e.s.u. to the same charge measured in e.m.u. This relation follows directly from the electromagnetic theory of light. By far the finest piece of work on this subject is that in 1906 by Rosa and Dorsey (1907). Their work was discussed in G.C. 1929 (pp. 9–10), and the discussion will not, in general, be repeated. Several small revisions are, however, needed. Their actual average result* is $2{\cdot}99712 \times 10^{10}$ cm.$^{1/2}$ sec.$^{-1/2}$ int.-ohm$^{1/2}$, rather than the rounded 2·9971 which they use. Furthermore, the best value of p, as stated in Section E ahead, is now $1{\cdot}00048 \pm 0{\cdot}00002$, in place of the value $1{\cdot}00051 \pm 0{\cdot}00002$ used in 1929. Their final result thus becomes $(1{\cdot}00048)^{1/2} \times (2{\cdot}99712 \times 10^{10}) = 2{\cdot}99784 \times 10^{10}$ cm./sec., or 299784 km./sec., for the value of c.† I again take 10 km./sec. as the probable error, and I again wish to emphasize that this work by Rosa and Dorsey is one of the most beautifully executed pieces of precision research in the entire history of science, and constituted certainly the most precise determination of c up till the last few years.‡

We have now obtained eight observed values of c, and these are the *only* data

* In specifying the unit on page 10, and again on page 60, of G.C. 1929, " int.-ohm$^{-1/2}$ " was inadvertently printed, in place of the correct " int.-ohm$^{1/2}$."

† By starting with their actual data for condensers in air (Rosa and Dorsey, 1907, p. 601), and using a slightly more accurate value of the dielectric constant of air, I get for the final reduced value 299783·2 km./sec., but it has not seemed worth while to make the trifling, but extensive, revisions that would be required bv this small change in the adopted value.

‡ Actually all that Rosa and Dorsey could measure, as already stated, was c in cm.$^{1/2}$ sec.$^{-1/2}$ int.-ohm$^{1/2}$ units, and until p had been correctly evaluated—which was not done until many years later—no final reliable value of c in cm.sec.$^{-1}$ units could be deduced. But this fact does not detract from the ***precision*** of their work.

worthy of consideration in obtaining a final most probable value. But in order to consider de Bray's hypothesis of the time variation of c, I list in table 1 also the five earlier " final declared values ", as given by de Bray (1927). Merely for the sake of argument I adopt also the stated probable error of the investigator himself, as given by de Bray, although as a result of the evidence uncovered in the recent investigations, it is possible that all of the assumed probable errors of these five earlier investigations are too small. We thus have 13 values of c. Table 1 gives also the mean epoch at which the work was performed and the actual published result. Another column gives the method employed, and another the adopted weight, taken, as usual, proportional to the reciprocal of the square of the probable error.

Table 1. Velocity of light

Author	Method	Epoch	Corrected result	Adopted probable error r	Adopted weight $100/r^2$	Original published result
Cornu-Helmert	TW	1874·8	299990	200	0·0025	299990
Michelson	RM	1879·5	299910	50	0·0400	299910
Newcomb	RM	1882·7	299860	30	0·1111	299860
Michelson	RM	1882·8	299853	60	0·0278	299853
Perrotin	TW	1902·4	299901	84	0·0142	299901
Rosa-Dorsey	EU	1906·0	299784	10	1·000	299710
Mercier	WW	1923·0	299782	30	0·111	299700
Michelson	RM	1926·5	299798	15	0·444	299796
Mittelstaedt	KC	1928·0	299786	10	1·000	299778
Michelson, Pease and Pearson	RM	1932·5	299774	4	6·250	299774
Anderson	KC	1936·8	299771	10	1·000	299764
Hüttel	KC	1937·0	299771	10	1·000	299768
Anderson	KC	1940·0	299776	6	2·778	299776

TW = toothed wheel ; RM = rotating mirror ; EU = electric units ;
WW = waves on wires ; KC = Kerr cell.

Using the data of the table, one now obtains the following results. The weighted average of the eight recent observations is $299776{\cdot}4_{37} \pm 1{\cdot}4_{76}(\pm 2{\cdot}7_1)$ km./sec. In this, and in subsequent expressions, the first stated probable error is that obtained on the basis of external consistency (r_e), and the second (in parenthesis) that given by internal consistency (r_i) (Birge, 1932 a). The ratio r_e/r_i is 0·544; in other words, the actual consistency of the results is about twice as great as is to be expected on the basis of the assigned probable errors. One may interpret this as the result of pure chance, or as an indication that the assigned probable errors are too large. I am inclined to believe that it is mostly the latter. On the other hand it seems best always to accept the larger of the two errors, and in order to allow for still unknown systematic errors, I shall adopt 4 km./sec. as the probable error. The final average, which I here *adopt*, is then

$$c = 299776 \pm 4 \text{ km./sec.}$$

This value is identical with that advocated in 1934 (Birge, 1934 b), and because

it is a value that has been used rather widely since that time, it is fortunate that the most recent work furnishes no grounds for suggesting a change.

If the Michelson, 1926, result is given a probable error of 10 km./sec., which I now think is the *minimum* permissible, the new weighted average is 299777·28; if his result is entirely ignored, the average is 299775·7 km./sec. It seems to me, therefore, that the presumption of evidence is in favour of 299776 rather than 299777, and in any case, a change of only one km./sec. would be devoid of significance.

The weighted average of the five older results is $299873{\cdot}8_6 \pm 8{\cdot}6_8 (\pm 22{\cdot}6_1)$ km./sec., and the ratio r_e/r_i is 0·384. Thus these older results are entirely consistent *among themselves*, but their average is nearly 100 km./sec. greater than that given by the eight more recent results. The cause of the sudden change in the experimentally determined values of c, at the opening of the 20th century, might be an interesting subject for investigation, but I would hesitate to believe that this 100 km./sec. change is real.

Using *all thirteen* results we get, as the weighted average, $299777{\cdot}8_2 \pm 2{\cdot}5_{78}$ $(\pm 2{\cdot}6_{94})$ km./sec., and the ratio r_e/r_i is 0·957. This last figure has considerable significance in connection with de Bray's hypothesis. It shows that the assumption of a *constant* value of c (for the period 1875 to 1940) *is* consistent with all the experimental evidence, when the probable error of each result is taken into consideration. Now my own philosophy of science—a philosophy that has been stated most eloquently by H. Poincaré (1913)—is that the simplest satisfactory explanation of a given set of phenomena should always be chosen. If the assumption of a constant value of c satisfies the data, it is gratuitous to postulate a systematically changing value.

Another, and more precise, method of testing the hypothesis of a linear variation of c with time is to calculate, by least squares, the best straight line passing through the data, plotted with the observed values of c as ordinates and the epoch (D), given in table 1, as abscissæ. For all 13 points the result is

$$c = 299779{\cdot}18 \pm 1{\cdot}97 - (0{\cdot}901 \pm 0{\cdot}192)\,(D - 1930) \text{ km./sec.}$$

The ratio $r_e/r_i = 0{\cdot}723$, so that the internal probable errors are 2·72 and 0·265 respectively, and hence, by internal consistency, the slope is 3·4 times its probable error. Statistically, this is a barely significant result. Incidentally, the calculated rate of change is less than one km./sec. per year, in contrast to the four km./sec. per year postulated by de Bray.

If, now, we use only the eight really reliable values, covering the period 1906 to 1940, the result is

$$c = 299777{\cdot}27 \pm 1{\cdot}37 - (0{\cdot}381 \pm 0{\cdot}158)(D - 1930) \text{ km./sec.}$$

and $r_e/r_i = 0{\cdot}490$. Hence the probable error of the slope, by internal consistency, is $\pm 0{\cdot}323$ km./sec. per year, and the calculated slope of 0·381 km./sec. per year has no significance whatever. This conclusion is fairly obvious from an inspection of table 1, but it is well to put it in quantitative form. I believe, then,

that there is now no experimental evidence of a linear variation in c, with time, although recent investigators have found continuously occurring irregular variations (from hour to hour, as well as day to day), due doubtless to a lack of complete stability in the experimental arrangements. Edmondson's (1934) suggested harmonic variation, as given by the equation *

$$c = 299885 + 115 \sin (2\pi/40)\,(D - 1901).$$

is already completely disproved, since it predicts $c = 299867$ km./sec. for 1940·0, contrasted with Anderson's observed $c = 299776 \pm 6$ km./sec. Thus, after a long and, at times, hectic history, the value of c has at last settled down into a fairly satisfactory " steady " state. The same thing can now be said of other important constants. Otherwise I would not be writing this paper !

SECTION B

THE NEWTONIAN CONSTANT OF GRAVITATION (G) AND THE MEAN DENSITY OF THE EARTH

In G.C. 1929 I adopted for G Heyl's final value at that time, namely, $G = (6{\cdot}664 \pm 0{\cdot}002) \times 10^{-8}$ dyne.cm.^{2} g.$^{-2}$ (Heyl, 1927). The adopted probable error is actually his average deviation from the average, for five determinations ranging from 6·667 to 6·661. The small masses used in the vibrating pendulum were made of platinum. Later Heyl continued his experiments, using gold and glass masses. His final averages (Heyl, 1930) for the three different materials differ considerably more than expected on the basis of their internal consistency, and he was not able to trace the origin of the discrepancies. Thus the final values for platinum, gold and glass are 6·664, 6·678 and 6·674 respectively, whereas the average deviations from the average for each material are only 0·002, 0·003 and 0·002 respectively. Heyl assigns weights of 3, 1 and 3, and thus obtains a weighted average of $6{\cdot}670_3 \pm 0{\cdot}002_7$, on the basis of external consistency. Because of the unexpected discrepancies between the three results, I increase the probable error to 0·005, and *adopt*

$$G = (6{\cdot}670 \pm 0{\cdot}005) \times 10^{-8} \text{ dyne.cm.}^2 \text{ g.}^{-2}$$

In G.C. 1929 I used $G \,.\, \delta$ (earth) $= 36{\cdot}797 \times 10^{-8}$ sec.$^{-2}$, where δ (earth) is the *average density* of the earth. The figure is taken from Henning and Jaeger (1926 b). With the present adopted value of G we get

$$\delta \text{ (earth)} = 5{\cdot}517 \pm 0{\cdot}004 \text{ g.cm.}^{-3}$$

SECTION C

RELATION OF THE LITRE (1000 ml.) TO THE CUBIC DECIMETER (1000 cm.^{3})

The litre is defined as the volume of a kilogram of air-free water at its maximum density. The kilogram is the mass of the " International Prototype Kilogram ", preserved at Paris. It has been customary to use " c.c." for the designation of

* This equation has been discussed (Birge, 1934 b).

litre/1000, but because of the frequent confusion with cm^3, to which it is ***not*** equal, the " Joint Committee for the Standardization of Scientific Glassware " recommended, in 1924, that the designation ***ml.*** (millilitre) be used in place of c.c. A very clear and detailed account of the entire situation has been given by Stott (1929). Unfortunately I had not been aware of the 1924 action until I read Stott's paper. Hence in G.C. 1929 I used the designation c.c. in place of ml.

As a result of the discovery of the isotopes of both hydrogen and oxygen, there is now a wide variety of different possible kinds of " water ". This situation, as it affects the litre, has been discussed by Riesenfeld and Chang (1936). Fortunately the density of " ordinary water " appears to be very uniform, although Riesenfeld and Chang say " it is now known that the litre determined from ocean water is 0·0015 cm^3 smaller than that from land water ". This difference is due to a change in the abundance ratio of the oxygen isotopes, a subject discussed in Section F.

The relation adopted in G.C. 1929 was, 1 litre $= 1000{\cdot}027 \pm 0{\cdot}001$ cm^3 Since then this figure has been corrected by Guillaume (1927) to 1000·028, and because of the variation just mentioned in regard to land water and ocean water, I increase its probable error to 0·002. Hence I ***adopt***

$$1 \text{ litre } (= 1000 \text{ ml.}) = 1000{\cdot}028 \pm 0{\cdot}002 \text{ cm}^3$$
$$= 1{\cdot}000028 \pm 0{\cdot}000002 \text{ dm}^3$$

The maximum density of water $\delta_m(H_2O)$ is accordingly

$$1/1{\cdot}000028 = 0{\cdot}999972 \pm 0{\cdot}000002 \text{ g.cm.}^{-3}$$

Since there is still some confusion on the matter, it may be well to emphasize that the litre and millilitre are ***not*** units of the present c.g.s. system, whose unit of volume is the ***cm^3***. They are units of volume in a system in which density (of water) replaces length as a fundamental unit.

SECTION D

THE NORMAL MOLE VOLUME (N.M.V.) OF AN IDEAL GAS

The normal mole volume of an ideal gas is the volume occupied by one gram mole at 0° c. under one atmosphere pressure. A brief account of the theory was given in G.C. 1929, and a much more detailed account later (Birge and Jenkins, 1934). The theory will not be repeated, but the basic equations will be given in order to make clear a number of changes in symbols.*

For an ideal gas we have

$$pV = mR_0T_0/M, \qquad \ldots\ldots(1)$$

where p = pressure, V = volume, m = mass in grams, M = gram molecular weight, T_0 = absolute temperature of the ice-point (0° c.), and R_0 = gas constant per mole. If $m/M = 1$ and $p = 1$, we get

$$V = V_0 = R_0T_0, \qquad \ldots\ldots(2)$$

* This is a subject in which there is no standard set of symbols. In particular, workers in the field of limiting gas densities often use quite different symbols from those usually employed in the general field of physical chemistry of gases.

where V_0 is the normal mole volume (N.M.V.) in the units chosen. Since the N.M.V. is thus merely the product R_0T_0, its dimensions by equation (1) are those of *pressure* × *volume*, for $m/M=1$ (one mole of gas).* There are consequently many units for N.M.V., depending on the units of pressure and of volume that are chosen. It is customary to choose one atmosphere as the unit of pressure, but the atmosphere (see Section G) may be in terms of *standard* gravity ($g_0=980{\cdot}665$ cm.sec^{-2}), or of 45° gravity ($g_{45}=980{\cdot}616$ cm.sec^{-2}). It is customary for workers in the field of limiting gas densities to express all experimental results in terms of g_{45}, and for convenience in quoting their work I shall do the same. The corresponding value of N.M.V. will be denoted V_{45}.

A real gas satisfies equation (1) only at zero pressure. Hence, from equations (1) and (2),

$$(pV)_0=mV_0/M, \qquad \ldots\ldots(3)$$

an equation holding for a real gas, where $(pV)_0$ is the value of pV at zero pressure Let us now *define*

$$(pV)_0/(pV)_1=1-\alpha, \qquad \ldots\ldots(4)$$

where $(pV)_1$ is the value of pV for one atmosphere pressure.† Let us also *define*

$$L=m/pV, \qquad L_1=m/(pV)_1, \qquad L_{\text{lim}}=m/(pV)_0. \qquad \ldots\ldots(5)$$

Since $m/V=\rho=$density, we have $L=\rho/p$ and $L=L_1=\rho$ for $p=1$. Thus L_1 is the actual density at one atmosphere, and L_{lim} is the density the gas would have, at one atmosphere, if it were an ideal gas. Both L and pV vary with p, for a real gas, and L_{lim} and $(pV)_0$ are the limiting values (obtained by extrapolation) at $p=o$.

From equations (4) and (5)

$$L_1/L_{\text{lim}}=(pV)_0/(pV)_1=1-\alpha. \qquad \ldots\ldots(6)$$

Then from equations (3), (5) and (6)

$$V_0 \text{ (or } V_{45})=M/L_{\text{lim}}=M(1-\alpha)/L_1 \qquad \ldots\ldots(7)$$

This last equation is used in the actual evaluation of V_0. The values of $(1-\alpha)$

* There is a more or less unavoidable ambiguity in the proper treatment of normal mole volume. If V_0 is considered as the actual volume per mole, its dimensions are necessarily those of volume. But the volume of a gas depends on the pressure, and when pressure is introduced, as in equation (1), we get V_0 *defined* by equation (2), from which its dimensions are necessarily those of *pressure* × *volume*, or *energy* (per mole). It merely happens that if energy is expressed in litre-atmos, or cm.^{3}-atmos units, then the *numerical* value of the energy per mole, for one atmosphere pressure, is just the *volume* (in litres or in cm.^{3}) per mole—i.e. the so-called normal mole volume. In G.C. 1929 I followed the example of the *International Critical Tables*, **1**, 17 (1926), and specified the unit of V_0 as volume per mole. But if V_0 ($=R_0T_0$) is to be used to calculate R_0 in various units, as is done in table *c*, at the end of the paper, then V_0 *must* be given the dimensions of energy. The actual calculations of R_0 in table *c* indicate the particular symbols to be used in equation (2), for certain specified units. Thus, from table *c*, $V_0=R_0'''T_0$ and $V_0'=R_0''T_0$. Also, from the specific definition of R_0 in table *c* (in contrast to R_0', R_0'' etc.), we get, in table *c*, $V_0A_0=R_0T_0$, as a measure of the actual pressure energy in *ergs* per mole, for an ideal gas at 0° c., under one atmosphere pressure (see Section G for A_0).

† The symbol $1+\lambda$ is often used in place of $1-\alpha$.

and L_1 (hence of L_{lim}) may be obtained in the same investigation. Sometimes $(1-\alpha)$, which by equation (4) requires a measurement only of the pV isothermals and not of the density, is measured in one investigation, and L_1 in another investigation. In order to evaluate V_0, oxygen is always used as the gas, since then $M=32$, without any experimental error. Knowing V_0, one can obtain the molecular weight M of any other gas from its observed L_{lim}.

It has been customary, in limiting gas-density work for the determination of molecular weights, to confine the observations to pressures of one atmosphere and less. For that range of pressure it is known that oxygen and other so-called permanent gases yield an essentially linear relation between L and p (and also between pv and p). But Cragoe (1941), in an exhaustive investigation, has shown that even in the case of a permanent gas, a much more reliable value of $1-\alpha$ can be obtained by the use also of data at higher pressures. For the range 0–100 atmospheres, Cragoe finds as the best least squares solution of the pV isothermals of oxygen

$$\frac{pV}{(pV)_1}=1-(95{\cdot}13\pm 0{\cdot}94)\times 10^{-6}\left(\frac{V_1}{V}-1\right)+2{\cdot}246\times 10^{-6}\left(\frac{V_1}{V}-1\right)^2, \quad \ldots\ldots(8)$$

where V_1 is the volume at one atmosphere pressure. If, now, we put $V=\infty$ (i. e., $p=0$) in equation (8), and use equation (6), we get*

$$(pV)_0/(pV)_1=1-\alpha=1{\cdot}0009535, \quad \ldots\ldots(9)$$

The stated uncertainty ($\pm 0{\cdot}94$) in equation (8) is not the least squares probable error, but a quantity about three times as large, designed to allow for systematic errors. I shall accept this estimate, and accordingly *adopt* $1-\alpha=1{\cdot}000953_5\pm 0{\cdot}000009_4$. In G.C. 1929 I adopted Baxter and Starkweather's least squares solution (Baxter and Starkweather, 1928) of their own data on oxygen (Baxter and Starkweather, 1926) at one atmosphere and less. This solution gave $1-\alpha=1{\cdot}000927\pm 0{\cdot}000030$, and $L_1=1{\cdot}428965\pm 0{\cdot}000030$ gram.litre^{-1} ($g=980{\cdot}616$).

Their *observed* value of L_1 is, however, $1{\cdot}42897\pm 0{\cdot}00003$, and if an *independent* value of $1-\alpha$, like Cragoe's, is to be adopted, only directly observed values of L_1 are relevant. I shall here *adopt* the value of L_1 (and its error) just quoted, and with the adopted $1-\alpha$ of Cragoe, we get $L_{\text{lim}}=L_1/(1-\alpha)=1{\cdot}427609\pm 0{\cdot}000037$, as contrasted with $L_{\text{lim}}=1{\cdot}427641\pm 0{\cdot}000055$ given by the G.C. 1929 values.

A proper consideration of all relevant data now available would require a very extended discussion. I shall, however, content myself with the following brief, categorical remarks. Batuecas (1935) has recalculated all of his data by least squares methods, as outlined and advocated by Birge and Jenkins (1934). His final result for oxygen is $L_{\text{lim}}=1{\cdot}42762\pm 0{\cdot}00005$. Moles, Toral and Escribano (1939), in a paper that is more accessible than most of the original references, obtain $L_{\text{lim}}=1{\cdot}427619$. I believe that these are the best values of L_{lim} that have been obtained since 1929, and they deviate from our present

* Cragoe incorrectly deduces $1-\alpha=1{\cdot}0009513$ from his equation.

adopted L_{litm} only by about one quarter of its probable error. They are also closer to it than to the value adopted in G.C. 1929.

Finally, with the adopted L_{litm} we get, for the units used,

$$V'_{45} = M/L_{\text{litm}} = 32/(1{\cdot}427609 \pm 0{\cdot}000037) =$$
$$22{\cdot}4151_0 \pm 0{\cdot}0006 \text{ litre.atmos.mole}^{-1} \ (g = 980{\cdot}616),$$

which I *adopt*. This may be contrasted with $22{\cdot}4146 \pm 0{\cdot}0008$ of G.C. 1929, and with the value 22·4149 given by the L_{litm} of either Batuecas (1935) or Moles *et al.* (1939).

To convert to cm^3 units, we multiply by 1000·028 (see Section C), and to convert to standard gravity (980·665) we multiply by the inverse ratio of the accelerations. Thus we get

$$V_{45} = (22{\cdot}4157 \pm 0{\cdot}0006) \times 10^3 \text{ cm.atmos.mole}^{-1} \ (g = 980{\cdot}616),$$
$$V'_0 = 22{\cdot}4140 \pm 0{\cdot}0006 \text{ litre.atmos.mole}^{-1} \ (g = 980{\cdot}665),$$
$$V_0 = (22{\cdot}4146 \pm 0{\cdot}0006) \times 10^3 \text{ cm}^3\text{atmos.mole}^{-1} \ (g = 980{\cdot}665).$$

It seems preferable to apply the expression " normal mole volume " to V_0 and V'_0, rather than V_{45} and V'_{45}.

SECTION E

RATIO OF INTERNATIONAL TO ABSOLUTE ELECTRICAL UNITS

All electrical measurements are still based on the international ampere, defined by means of the silver voltameter, and the international ohm, defined by the resistance of a column of mercury. It had been hoped to adopt, on January 1, 1940, the *absolute* electrical units as the official units. Effectively such action can be taken by merely accepting official values for p and q, where one int. ohm $= p$ abs. ohm, and one int. amp. $= q$ abs. amp.

In preparation for such action, all the national physical laboratories undertook, some years ago, to obtain new and more trustworthy values of p and q. Furthermore, there have now been carried out extensive intercomparisons of the standard resistances and standard cells of the various countries, and from such comparisons a " mean international ampere " and a " mean international ohm " have been defined. Any one laboratory can thus convert its own national standards into these international standards.

Because of the international situation, it has been impossible for all the laboratories to complete the work on p and q, and accordingly no attempt has yet been made to adopt official values. But both the National Bureau of Standards, Washington (N.B.S.), and the National Physical Laboratory (N.P.L.) have completed the necessary investigations, with results that are in beautiful agreement, although the value of q obtained by each laboratory is appreciably different from the value considered most reliable in 1929. The actual results are, very briefly, as follows:—

(1) Curtis, Curtis and Critchfield (1939) obtain $q = 0{\cdot}99986$ in terms of the N.B.S. int. amp., which becomes 0·999868 in terms of the mean int. amp.

(2) Vigoureux (1938 a) gets $q = 0{\cdot}99986$ in terms of the N.P.L. int. amp., which becomes 0·99985 in terms of the mean int. amp. In both of these investigations an attempt was made to get a result trustworthy to about 2 parts in 10^5. From these two results I get as an *adopted* value, $q = 0{\cdot}99986 \pm 0{\cdot}00002$, in terms of the mean int. amp.

In the case of p the results are as follows: (1) Curtis, Moon and Sparks (1938) obtain $p = 1{\cdot}000468$ in terms of the N.B.S. int. ohm, or 1·000472 in terms of the mean int. ohm; this result is the weighted average of the results of two investigations by these authors (1936, 38). (2) Two investigations at the N.P.L., one by Hartshorn and Astbury (1937), and the other by Vigoureux (1938 b), each give $p = 1{\cdot}00050$ in terms of the N.P.L. int. ohm, or 1·000508 in terms of the mean int. ohm. A slightly higher accuracy is claimed for the N.B.S. work, and I accordingly adopt as a final weighted average $p = 1{\cdot}00048 \pm 0{\cdot}00002$, in terms of the mean int. ohm.

This adopted value of p agrees fairly well with the value $1{\cdot}00051 \pm 0{\cdot}00002$ adopted in G.C. 1929. But the new value of $q = 0{\cdot}99986 \pm 0{\cdot}00002$ deviates considerably both from the value $0{\cdot}99995 \pm 0{\cdot}00005$ adopted in G.C. 1929, and $0{\cdot}99993 \pm 0{\cdot}00002$ adopted in 1938 (Birge, 1938). This change in the value of q makes many small changes in other constants, but further appreciable changes in either p or q are unlikely to occur for some time.

The relations between the international and absolute units of other electrical and magnetic quantities are listed in table *b*, at the end of this paper.

SECTION F
THE ATOMIC WEIGHTS OF CERTAIN ELEMENTS

The situation in regard to the atomic weights of many elements has been completely changed since G.C. 1929 was written. In the first place, the discovery of isotopes of oxygen (^{17}O and ^{18}O) has made possible two different scales of atomic weights: (1) the physical scale, based on $^{16}O = 16{\cdot}0000$, and (2), the chemical scale, based on the ordinary mixture of oxygen isotopes $O = 16{\cdot}0000$. In the second place, atomic weights of many individual isotopes, on the physical scale, can now be measured by means of the mass spectrograph with an accuracy equalling, or even exceeding, any chemical determination of atomic weight. Hence, in the case of certain elements such as H, He, C and N, the *most accurate* determination of the atomic weight, on the chemical scale, of the ordinary mixture of isotopes (in other words, the ordinary atomic weight of chemistry) can be made by starting with the mass spectrograph values of the individual isotopes.

In order to carry out such a calculation one must know (1) the mass of each isotope, with its probable error, on the physical scale, as determined with the mass spectrograph; (2) the relative abundance, with its probable error, of the various isotopes; (3) the factor r (with its probable error) by which one converts from the physical to the chemical scale of atomic weights.*

* See R. T. Birge (1939 a) for the explicit formula needed to calculate, by propagation of errors, the final probable error of the specified mixture of isotopes.

The value of r is obtained from the observed masses and relative abundance of the oxygen isotopes, and the value here adopted is $r = 1{\cdot}000272 \pm 0{\cdot}000005$. This value results from the following assumed isotopic abundance,

$$^{16}O : {}^{18}O : {}^{17}O = (506 \pm 10) : 1 : (0{\cdot}204 \pm 0{\cdot}008),$$

together with the masses $^{16}O = 16{\cdot}0000$ (by definition), $^{18}O = 18{\cdot}0049$ and $^{17}O = 17{\cdot}0045$ (Hahn, Flügge and Mattauch, 1940). Whatever small probable errors may exist for the rare isotopes ^{18}O and ^{17}O have a completely negligible effect on the resulting probable error of r.

The ratio $^{17}O/^{16}O = 0{\cdot}204 \pm 0{\cdot}008$ is due to Murphey (1941). The important ratio $^{16}O/^{18}O$ presents more of a problem. There are three rather recent determinations. Manian, Urey and Bleakney (1934) obtain 514 ± 13 as a final weighted average of observations on oxygen derived from five different sources (meteorites and rocks). The ratios obtained from different sources are all, internally, quite consistent, but vary from 493 to 540 for the different sources. These results, however, are derived by one method. With another method the 493 ratio changes to 526, and the 540 ratio becomes 515, etc. Hence the variations may be only experimental in origin.

The other two investigations are those by Smythe (1934), who obtained a ratio 503 ± 10 for oxygen derived from PbO_2 and by Murphey (1941), who obtained 500 ± 15 from various samples of commercial oxygen and of air. There is probably little difference in the accuracy of these three investigations, and I therefore adopt the unweighted average 506. In view of the variations in the experimental values just quoted, and the additional evidence now to be presented, it would appear that one must take the probable error at least as large as ± 10.

Apart from the question of the *absolute* value of the $^{16}O/^{18}O$ ratio, we have the very interesting question of the possible real variation of the ratio in the case of oxygen from various sources. If such a variation exists, there must necessarily be a corresponding variation in the value of r, depending on the source of the oxygen used. Extensive and extremely precise work on this subject has been carried out in the past six years. The method is to use oxygen in the form of water, and to measure the *relative density* of the water prepared with oxygen from various sources. The density of the water is stated in terms of " gamma units ", where one gamma is the *excess* density of the water in question, in parts per million, over ordinary water. It is possible at present to obtain an accuracy of one- or two-tenths of a gamma unit.

Now if the changing density of the water is due solely to a change in the $^{16}O/^{18}O$ ratio of the oxygen, one can easily show that an *increase* of one gamma unit in the density of water equals a *decrease* of 2·04 units in the $^{16}O/^{18}O$ ratio, based on 506 as the standard value. Thus *variations* as small as 0·1 per cent in the $^{16}O/^{18}O$ ratio can be detected in this way. It is not practical to go into details here, but a comprehensive picture of the entire experimental situation can be gained from three recent papers, by Greene and Voskuyl (1939), Swartout

and Dole (1939) and Dole and Slobod (1940). The facts, briefly stated, are as follows.

"Ordinary water" (from fresh-water lakes as well as the tap-water of various cities) shows an exceedingly uniform density. Water from the ocean (after distillation) has a density one to two gamma units in excess of ordinary water.* The water prepared from the oxygen in geologically old iron ores does not differ significantly from ordinary water. But water from the oxygen of carbonate rocks of all kinds is heavier by about eight gamma units. Furthermore, it is known that in the case of CO_2 in contact with water, there is an excess of ^{18}O in the CO_2, equivalent to about ten gamma units. Finally, water prepared from atmospheric oxygen is heavier than ordinary water by 6·6 gamma units. This last difference is the so-called "Dole Effect", and the 6·6 gamma unit value is due to Swartout and Dole (1939), whose investigation is the latest and most accurate on this subject.

We can now draw the following conclusions. The oxygen in most rocks, regardless of geologic age, has a very uniform isotopic composition, which is the same as that of oxygen in ordinary water. But carbonates, which come from animal life, have a distinctly higher ^{18}O abundance, and the oxygen of the atmosphere, which is believed to come almost wholly from the photosynthesis of plants, has a similar higher abundance. In both cases the differential effect is presumably due to the natural higher abundance of ^{18}O in CO_2, which has been explained on theoretical grounds. From the foregoing evidence we can also feel quite certain that the variations in the isotopic abundance found by Manian, Urey and Bleakney (1934) are almost entirely experimental in origin.

It should be noted, in this connection, that the work just cited on the density of water also leads to the abundance ratio $^1H/^2H = 6900 \pm 100$, given in table *a'*. The involved experimental procedures designed to isolate a possible changing $^1H/^2H$ ratio from a possible changing $^{16}O/^{18}O$ ratio are described in detail by Swartout and Dole (1939). Furthermore, if we now assume that the *adopted* $^{16}O/^{18}O$ ratio $= 506$ applies to the oxygen of ordinary water, and hence to that in most rocks (*excluding* carbonates), then the ratio for atmospheric oxygen is about 492.†

All of the atomic weight values that have been adopted for the purposes of the present investigation are listed in table *a'*, at the end of the paper. No real details will be given here. The adopted values of H, He, C and N on both the physical and chemical scales are based on the mass spectrograph values of the physical scale, with the indicated isotopic abundance. In all cases I have used what appears to be the best average of recent determinations. In the case of nitrogen there is given also the present average result $(14{\cdot}0086 \pm 0{\cdot}0007)$ derived from limiting gas-density investigations.

* It is this excess that is mentioned in connection with the litre, in Section C.

† A gamma value of 6·6 equals a change of 13·5 in the $^{16}O/^{18}O$ ratio.

One of the best determined of all chemical ratios, based on gravimetric work, is $AgNO_3/Ag = 1{\cdot}574790$, which appears to have a probable error of 0·000005 or less.* From $N = 14{\cdot}00740 \pm 0{\cdot}00012$, as determined from mass spectrograph work (see table *a'*), one gets immediately $Ag = 107{\cdot}8783_5 \pm 0{\cdot}0009_5$, and this result is the basis for my remarks about the atomic weight of silver, in the introduction to this paper. From the limiting density value $N = 14{\cdot}0086 \pm 0{\cdot}0007$, and the same $AgNO_3/Ag$ ratio, one gets $Ag = 107{\cdot}880_2 \pm 0{\cdot}001_5$. As already noted, I *adopt* $Ag = 107{\cdot}880 \pm 0{\cdot}002$, pending a more comprehensive examination of this important question.

The value for iodine ($126{\cdot}915 \pm 0{\cdot}004$) results from a study of the latest data, much of which is discussed by Baxter.† The present adopted value deviates considerably from the value ($126{\cdot}932 \pm 0{\cdot}002$) adopted in G.C. 1929, but we now have almost identical results yielded by several quite different experimental methods, and except for the remaining uncertainty due to its dependence on the value of Ag, the atomic weight of iodine seems well established.

The atomic weights of calcium, chlorine and sodium are based on chemical determinations. The first is needed in the evaluation of the electronic charge. The last two are fundamental atomic weights, used in some of the results for iodine and for many other elements.

SECTION G

THE STANDARD ATMOSPHERE (A_s)

The standard atmosphere is defined as the pressure exerted by a column of mercury 76 cm. high, of standard density (i.e., at 0° c., A_0), under standard gravity (g_0). Although A_0 is itself used in the definition, the very slight variation of the density of mercury with pressure makes this inclusion of A_0 possible. The standard atmosphere is then defined by

$A_0 = H_0 \cdot \rho_0(\mathrm{Hg}) \cdot \delta_m(H_2O) \cdot g_0$, where $H_0 = 76{\cdot}000$ cm. = height of standard barometer;

ρ_0 = standard specific gravity of Hg (at 0° c., A_0), referred to air-free water of maximum density;

$\delta_m(H_2O)$ = maximum density of water = $0{\cdot}999972 \pm 0{\cdot}000002$ g.cm.$^{-3}$ (Section C);

g_0 = standard gravity = 980·665 cm.sec.$^{-2}$ (Section B).

The only quantity that remains to be evaluated is thus ρ_0. In G.C. 1929 I adopted $\rho_0 = 13{\cdot}59546$, which is the unweighted average of four determinations listed by Henning and Jaeger (1926 a). The discussion by these authors indicates, however, that only the last investigation of the four—that by Scheel and Blankenstein (1925)—is worthy of serious consideration, and even their result of 13·59549 is open to some question, for the following reason. Their work, carried out by the hydrostatic method, gave two results. One result, 13·59540, was obtained with mercury " E ", a quantity of mercury used at the *Reichsanstalt* in standard electrical resistances. It had often been distilled. The other result, 13·59558,

was obtained with mercury " III ", a quantity of mercury used in the measurement of small pressures. For the present investigation it was distilled in vacuum and a trifling amount of impurity was found. Since this second sample had seldom been distilled, the difference of 13 parts in 10^6 might be real, due to some difference in isotopic composition brought about by the distillation. But Jaeger and Steinwehr (1926), using the same two samples of mercury, found later, by direct observation of *relative* density (a more accurate process), that the two densities agreed within a very few parts in 10^6. Hence the previously found difference must be purely experimental, and under the circumstances one can only adopt the average value, with half the difference as its probable error, i.e., $13{\cdot}59549 \pm 0{\cdot}00009$.

Meanwhile a new investigation has been carried out by Batuecas and Casedo (1936). They used mercury from the Almaden mines, and with the pycnometer method obtained $\rho_0 = 13{\cdot}59539$. Their published probable error is only 0·00001, but a study of their work indicates that 0·00006 is a more reasonable estimate. With this latter figure, the weighted average of the two investigations is $\rho_0 = 13{\cdot}59542 \pm 0{\cdot}00003$, ($\pm 0{\cdot}00005_0$). As usual I adopt the larger of the two probable errors, in this case that based on internal consistency.

Thus D_0, the density of mercury at 0°C., A_0 ($= \rho_0 \cdot \delta_m$), is $D_0 = 13{\cdot}59504_0 \pm 0{\cdot}00005_7$ g.cm.$^{-3}$, and finally *

$$A_0 = D_0 \times 76 \times 980{\cdot}665 = (1{\cdot}013246 \pm 0{\cdot}000004) \times 10^6 \text{ dyne.cm.}^{-2}$$

or

$$A_{45} = D_0 \times 76 \times 980{\cdot}616 = (1{\cdot}013195 \pm 0{\cdot}000004) \times 10^6 \text{ dyne.cm.}^{-2}$$

SECTION H

THE ABSOLUTE TEMPERATURE OF THE ICE-POINT (T_0)

In G.C. 1929 the adopted value of the ice-point was $T_0 = 273{\cdot}18 \pm 0{\cdot}03°$ K. This result was derived from Roebuck's observed value 273·15° K. and Henning and Heuse's value † 273·19° K. Since then very extensive work has been carried out, and fortunately this work has been thoroughly reviewed in two very recent summary papers. Hence all details may be omitted here. Both of the papers in question were read at the recent Symposium on Temperature, organized by the American Institute of Physics. Roebuck and Murrill (1941) give an extensive table of determinations of T_0, with an average value of 273·170° K. They advocate $273{\cdot}17 \pm 0{\cdot}02°$ K. as the present best value. Beattie (1941) gives a similar table with an average of 273·165° K. He also describes his own recent work, which is the latest and possibly the most accurate that has been done in this field. Two different methods of handling his own data lead to 273·165° K. and 273·167° K., respectively.

* Had it not been for my premature rounding of data, the value $A_0 = 1{\cdot}013249 \times 10^6$ given in G.C. 1929 would have been 1·013250, and hence the present adopted A_0 and A_{45} values are each smaller by 4 dyne.cm.$^{-2}$

† See G.C. 1929, pp. 28–29, for all references.

Meanwhile W. M. Keesom and W. Tuyn, acting as a sub-committee for the *Institut International du Froid*, issued in 1937 a mimeographed report in which they advocate 273·15° K. as the best value of T_0.* It may be noted, in this connection, that the Leiden value of 273·144° K. is definitely lower than most recent results. So far as I am aware, this report has never been acted on, but according to *Nature*,† one of the resolutions which was to be presented at a projected, but never held, meeting of the *Institut International du Froid* was that $T_0 = 273{\cdot}15 \pm 0{\cdot}02°$ K. I believe that Keesom and Tuyn did not have available Beattie's recent result of 273·165° or 273·167° K.

In view of the foregoing evidence, it appears that the best value is now

$$T_0 = 273{\cdot}16 \pm 0{\cdot}01° \text{ K.},$$

and Dr. F. G. Brickwedde informs me that this is the opinion of the National Bureau of Standards. It would also appear that the value of the ice-point is now well established and that one can, with some confidence, adopt $\pm 0{\cdot}01°$ K. as its probable error.

SECTION I

THE JOULE EQUIVALENT (J) AND THE ELECTRICAL JOULE EQUIVALENT (J')

At the time G.C. 1929 was written there was available one outstanding piece of work on the electrical Joule equivalent and one on the mechanical Joule equivalent. Since then there has appeared another, and even more precise, evaluation of the electrical equivalent. Hence there are three results from which to get a weighted average.

The older work on the electrical equivalent is that by Jaeger and Steinwehr (1921; referred to in the rest of this section as "J. and S.") They obtained data over the range 5° to 50° C. As stated briefly in G.C. 1929, their published second degree power series equation, from which their adopted value of $J'_{15} = 4{\cdot}18420$ int. joules is derived, does not properly represent their data. In particular, the residuals show a serious trend in the vicinity of 15° C., just where it is most important that an accurate representation of the data be secured. I found, however, that a very satisfactory fit could be secured by means of a fourth degree power series, and from the new equation I obtained $J'_{15} = 4{\cdot}1832_7$ int. joules, with an estimated probable error of $\pm 0{\cdot}0008_0$ int. joule. This correction by the writer has been generally ignored in later German papers, although there is not the slightest question as to its validity.

Quite recently Osborne, Stimson and Ginnings (1939; referred to in the rest of this section as "O.S.G.") have made a new and extremely careful determination of J'_{15}. In their paper they discuss critically all of the older work. They too ignore the G.C. 1929 correction to th J. and S. result, but like the

* See also the lecture by H. B. G. Casimir (1939), in whic $T_0 = 273{\cdot}15°$ K. is advocated.

† *Nature, Lond.*, **145**, 597 (1940).

writer, find a trend in the J. and S. residuals. This trend they proceed to correct in a quite different way, but their corrected result ($J'_{15}=4{\cdot}1832$) is almost identical with my own. Their own experimental results, which cover the range 0° to 100° c., are given in the form of a table, in which J_{15} is listed as 4·1850 int. joules. Their actual result at 15° c. is, however, 4·18502.* No explicit estimate of their probable error is given, but Dr. Osborne writes†: "We suspect no sources of systematic errors which, if removed, would change our result by more than one part in 10,000, although there may be an even chance that they were present." This is a field in which systematic errors are important, and I take ±0·0004 (i. e., one part in 10,000) as the probable error of the O.S.G. result.

The direct measurement of the mechanical equivalent is that by Laby and Hercus (1927; referred to in the rest of this section as "L. and H."). This work also was recalculated by the writer, in G.C. 1929, and certain data were discarded. The value, as thus recalculated, is $J_{15}=4{\cdot}1852_6\pm0{\cdot}0008_0$ abs. joules, and this result was later accepted by Laby and Hercus (1935). Since the average residual in the L. and H. work is somewhat smaller than in the J. and S. work, I am now increasing the J. and S. probable error to 0·0009, and decreasing that of L. and H. to 0·0007. In order to use also the value of $J_{15}=4{\cdot}1852_6\pm0{\cdot}0007_0$ abs. joules, it must first be reduced to int. joules by means of the factor† $pq^2=1{\cdot}00020\pm0{\cdot}00004_5$. The result is $J_{15}=4{\cdot}1844_2\pm0{\cdot}0007_0$ int. joules.

We now have three values of J_{15}, namely: (J. and S.) $4{\cdot}1832_7\pm0{\cdot}0009_0$, (O.S.G.) $4{\cdot}1850_2\pm0{\cdot}0004_0$, and (L. and H.) $4{\cdot}1844_2\pm0{\cdot}0007_0$. The weighted average is $J_{15}=4{\cdot}1846_{64}\pm0{\cdot}0002_{81}(0{\cdot}0003_{24})$. The r_e/r_i ratio is 0·867, an entirely satisfactory value. I therefore *adopt*, for the electrical Joule equivalent, at 15° c.,

$$J_{15}=4{\cdot}1847\pm0{\cdot}0003 \text{ int.-joule.cal}_{15}^{-1}$$

Multiplying by pq^2, we get for the *adopted* (mechanical) Joule equivalent, at 15° c.,

$$J_{15}=4{\cdot}1855\pm0{\cdot}0004 \text{ abs.-joule.cal}_{15}^{-1}.$$

The adopted value of J'_{15} is 0·0012 larger than that given in G.C.1929, but J_{15} is only 0·0003 larger.

SECTION J

THE FARADAY (*F*)

The Faraday F is defined as the quantity of electricity carried in electrolysis by one gram equivalent M of any atom or molecule. Hence, if a charge Q is associated with m grams mass, we have

$$F=Q/(m/M) \qquad \ldots\ldots(1)$$

and $F=Q$ when $m/M=1$ (one gram equivalent). The number of atoms or molecules in a gram equivalent is N_0/V and each carries a charge Ve, where V

* See table *b*. In G.C. 1929, $pq^2=1{\cdot}00041\pm0{\cdot}00010$ was used.

† Private communication from Dr. Osborne.

is the valence. Hence the total charge carried by one gram equivalent is $F = N_0 \epsilon$, where N_0 = Avogadro number and ϵ = electronic charge. This last equation is used, in Section K, for the evaluation of ϵ, from the measured values of F and N_0.

Equation (1) may also be written

$$F = M/(m/Q) = M/E, \qquad \ldots\ldots(2)$$

where $E(=m/Q)$ is the electrochemical equivalent, i.e., the mass associated with unit charge. Equation (2) is used for the actual evaluation of F, and if the substance is univalent, the gram equivalent M is merely the atomic or molecular weight in grams. E is measured by an electrolytic cell, and, unfortunately, the mass E^* actually deposited or liberated per unit electric charge in such a cell is not necessarily the true mass E associated with unit charge. Let us denote by F^* the resulting *apparent* value of F, as obtained from $F^* = M/E^*$.

In order to obtain a trustworthy value of F, the first requirement is to select an electrolytic reaction that is likely to yield a value of E^* as close as possible to the true E. Only two such reactions have thus far been seriously proposed. These are the reactions occurring at the cathode of the silver voltameter, and at either the anode or the cathode of the iodine voltameter. Unfortunately the present resulting values of F^* given by these two voltameters differ by an amount greater than had been expected from the obvious experimental errors. Hence there has been, for many years, considerable uncertainty as to the true value of F, and considerable controversy as to the best value to adopt, on the basis of the available evidence.

A brief discussion of the situation, as it existed in 1929, is given on pp. 33–36 of G.C. 1929. At that time I recommended a value of F based wholly on the silver voltameter. At the present time I shall recommend a certain weighted average of the silver and iodine voltameter results. This is a subject that for more than one reason demands a detailed discussion, and I have, in fact, already prepared such a discussion. There is, however, no space for it here, and I accordingly content myself with the following brief statement.

Since both silver and iodine are univalent elements, equation (2) becomes, for these substances,

$$F^*_{Ag} = Ag/E^*_{Ag} \qquad \ldots\ldots(3), \qquad\qquad F^*_I = I/E^*_I, \qquad \ldots\ldots(4)$$

where Ag and I represent the atomic weights in grams (given in Section F), and E^*_{Ag} and E^*_I the apparent electrochemical equivalents in grams per coulomb, as measured by the silver and iodine voltameters. Then, from equation (3),

$$F^*_{Ag} = \frac{107{\cdot}880 \pm 0{\cdot}002}{0{\cdot}00111800} = 96493{\cdot}7 \pm 1{\cdot}8 \text{ int. coulombs}, \qquad \ldots\ldots(5)$$

since $E^*_{Ag} = 0{\cdot}00111800$ gram (with *no* error) by the *definition* of the international coulomb.

In the case of the iodine voltameter it is necessary to measure not only the mass of iodine released, but also the quantity of electricity Q that has passed

through the cell. To measure this latter quantity, Vinal and Bates (1914), who have carried out the only really precise determination, attached a silver voltameter in series to the iodine voltameter, so that the same current passed through both cells. They also, in the case of six runs, measured the current by means of a standard cell and standard resistance. The two values of Q thus obtained differ for the six runs by a maximum of 4 parts in 10^4, with an algebraic average difference of only 1·5 parts in 10^5.

Thus Q, in int. coulombs, is measured by m_{Ag}/E^*_{Ag}, where m_{Ag} is the measured mass of silver deposited in the silver voltameter. Then, from its definition, $E^*_I = m_I/Q$, where m_I is the corresponding mass of iodine released at the anode of the iodine voltameter. Hence we get

$$E^*_I = \frac{m_I}{(m_{Ag}/E^*_{Ag})} = \frac{E^*_{Ag}}{(m_{Ag}/m_I)} \qquad \ldots\ldots(6)$$

Vinal and Bates' experimental result is $m_{Ag}/m_I = 0{\cdot}850173_4 \pm 0{\cdot}000016$,† and from equation (6)

$$E^*_I = \frac{0{\cdot}00111800}{0{\cdot}850173_4 \pm 0\ 000016} = (131{\cdot}5026 \pm 0{\cdot}0025) \times 10^{-5}\,\text{g}. \qquad \ldots\ldots(7)$$

Finally, from equation (4),

$$F^*_I = \frac{126{\cdot}915 \pm 0\ 004}{(131{\cdot}5026 \pm 0\ 0025) \times 10^{-5}} = 96511{\cdot}4 \pm 3{\cdot}4_6 \text{ int. coulombs.} \qquad \ldots\ldots(8)$$

The difference between the results of equations (5) and (8) is $17{\cdot}7 \pm 3{\cdot}9_0$ int. coulombs, and it is this serious discrepancy (4·5 times its probable error) that has caused so much confusion and controversy.

My present solution of the difficulty, with all detailed explanation omitted, is as follows. It is known (Vinal and Bouvard, 1916) that inclusions exist in the silver deposit, to the average amount of 4 parts in 10^5. Correction for such inclusions raises F^*_{Ag} by 3·9 coulombs to 96497·6 int. coulombs. Three different types of systematic error in the iodine voltameter are listed by Vinal and Bates (1914). *Each* of these errors tends to make the observed F^*_I too high, and the discussion of Vinal and Bates indicates that these three errors may well aggregate 5 parts in 10^5, or 4·8 coulombs. With this assumption F^*_I becomes 96506·6 int. coulombs.

There are, however, various other possible sources of systematic error in *both* voltameters. Most of those mentioned for the silver voltameter tend to make F^*_{Ag} too *high*, and that is just the reason why, in 1929, I did not make the adopted value of F still higher, by using some weighted average of F^*_{Ag} and F^*_I. But Dr. J. W. M. DuMond has recently suggested‡ that there may be a preferential deposition of the two isotopes of silver, in the silver voltameter.

† This is my own calculated weighted average of their ten runs, with the probable error based on external consistency.

‡ Private communication.

The sign of this error, if it exists, is not known. Since the remaining types of possible systematic error in the iodine voltameter are partly positive and partly negative, I think it not unreasonable to include a probable error of 4 parts in 10^6, in the case of each of the voltameters, to cover various additional sources of errors. On combining the new probable errors with those already given, we obtain

$$F_{Ag}^{*} = 96497{\cdot}6 \pm 4{\cdot}25 \text{ int. coulombs}, \quad F_{I}^{*} = 96506{\cdot}6 \pm 5{\cdot}18 \text{ int. coulombs}.$$

The discrepancy thus becomes $9{\cdot}0 \pm 6{\cdot}7_0$ int. coulombs, which is not unreasonable. Hence it is now legitimate to weight the two results according to their probable errors, and we thus obtain $F^{*} = 96501{\cdot}22 \pm 2{\cdot}98$ ($\pm 3{\cdot}29$) int. coulombs, which we consider as our most probable value of the true F.

The probable errors just given make no allowance for the uncertainty in the deliberate corrections that have been made to the values of F, and because of the numerous possible sources of error, both known and unknown, it seems best to adopt a considerably larger probable error for the final weighted average. For some years past I have suggested a probable error of 10 coulombs, and I renew the suggestion. Hence I *adopt*, for the value of F on the *chemical scale*,

$$F = 96501{\cdot}_2 \pm 10 \text{ int. coulombs}.$$

Then, from the adopted value of q (Section E)

$$F = 96487{\cdot}_7 \pm 10 \text{ abs. coulombs}$$

$$= 9648{\cdot}7_7 \pm 1{\cdot}0 \text{ abs. e.m.u.},$$

$$F' = Fc = (2{\cdot}89247 \pm 0{\cdot}00030) \times 10^{14} \text{ abs. e.s.u.}$$

Finally, from the adopted value of r (Section F), we obtain, for the values of F on the *physical scale*,

$$F = 96514{\cdot}_0 \pm 10 \text{ abs. coulombs}$$

$$= 9651{\cdot}4_0 \pm 1{\cdot}0 \text{ abs. e.m.u.},$$

$$F' = Fc = (2{\cdot}89326 \pm 0{\cdot}00030) \times 10^{14} \text{ abs. e.s.u.}$$

The adopted value of $96501{\cdot}_2$ int. coulombs, on the chemical scale, is 7·2 coulombs higher than the F_{Ag}^{*} value 96494, which I have advocated from1 929 up till now. On the other hand, the final reduced value $F = 96514{\cdot}_0$ abs. coulombs, on the physical scale, is very close to the value $96513{\cdot}_1$ that I used in 1938 (Birge, 1938). The close coincidence is due to two nearly compensating changes: (1) the change of q from 0·99993 to 0·99986, and (2) the change of F, in int. coulombs, from 96494 to $96501{\cdot}_2$.

On the basis of the adopted value of F, the *true* electrochemical equivalents are, by equation (2),

$$E_{Ag} = (1{\cdot}11807 \pm 0{\cdot}00012) \times 10^{-3} \text{ g.abs.-coulombs}^{-1} \text{ (chemical scale), and}$$

$$E_{I} = (1{\cdot}31535 = 0{\cdot}00014) \times 10^{-3} \text{ g.abs.-coulombs}^{-1} \text{ (chemical scale).}$$

SECTION K

THE AVOGADRO NUMBER (N_0) AND THE ELECTRONIC CHARGE (e)

We come now, in conclusion, to the three atomic constants, e, m and h.* One or more of these constants enters into nearly every formula of modern physics and, because of the resulting theoretical relations, it is possible to evaluate many different combinations of the three constants. In fact, in the case of m and h, it is possible to measure them with any real precision *only* in certain combinations, and the three simplest quantities that can thus be determined experimentally are e, e/m and h/e. Hence the three remaining sections of the paper will be devoted to these three quantities, respectively.

As is doubtless well known to the reader, the various measured functions of e, m and h have hitherto exhibited certain serious inconsistencies. This situation has been directly responsible for a flood of experimental investigation, with an accompanying flood of critical discussion. In fact, the major portion of all experimental work on the general physical constants, during the past decade, has been concerned with the atomic constants. Accordingly the major portion of any detailed paper on the general physical constants should be devoted to the atomic constants, and since I have an especial interest in such constants, there is all the more reason for me to consider them at length.

Actually it is necessary to take the opposite course. As already noted, there have been numerous critical papers on this subject, from which the reader can get full details. Among the more recent papers are those by Wensel (1939), Dunnington (1939), DuMond (1939, 40) and Kirchner (1939). Wensel is concerned primarily with the radiation constants σ and c_2, and he treats the atomic constants only in connection with the evaluation of the radiation constants. Much the longest paper, and the only one that gives any experimental details, is that by Kirchner. Dunnington treats the three atomic constants by means of what he and others have termed the Birge-Bond diagram (Bond, 1930, 31; Birge, 1932 b). DuMond treats them in terms of a new type of diagram. As I have pointed out (Birge, 1940 a), there is an infinite number of such possible diagrams, and since each of them attempts to represent in two dimensions the relations between three quantities (e, m and h), each one suffers from the same limitations. Hence the choice of the best form of diagram is more or less a matter of taste.

The history of the values of e, m and h is really a most curious one. In G.C. 1929 I adopted what was known as "Millikan's value of e" ($\sim 4{\cdot}77 \times 10^{-10}$ e.s.u.). Even at that time a quite different value of e ($\sim 4{\cdot}80 \times 10^{-10}$ e.s.u.) had just been obtained by means of the wave-lengths of x rays, as measured with ruled gratings. Three years later (Birge, 1932 b), I still felt that no appreciable modification of the value of e was called for, and this opinion was based not on any special trust in Millikan's work, but rather on the fact that the functions of e, m and h measured with the greatest apparent accuracy were much more consistent with the "oil-drop" (Millikan) value of e than with the "ruled-grating" value.

* e = charge on the electron, m = mass of the electron for vanishingly small velocity, and h = Planck's constant.

Meanwhile extensive experimental and theoretical work has now made it quite clear that the ruled-grating value of e must be essentially correct, and repeated new measurements of the viscosity of air have shown that the value used by Millikan was seriously in error. The present most reliable values of the viscosity raise the oil-drop value of e not only up to the grating value, but even above it. But at the moment the grating value appears to have much the greater accuracy, and it alone will be evaluated here.

As a result of the change in the adopted value of e, a number of new and apparently serious discrepancies arose. I immediately pointed out (Birge, 1935) the gravity of the new situation, and a few months later I noted (Birge, 1936) that all of the important remaining discrepancies could be reduced to the statement that the directly measured values of h/e differed seriously from the value calculated indirectly by means of the observed values of e and e/m and the Bohr formula for the Rydberg constant. When I started to write this paper, I believed that this final remaining discrepancy had at last been resolved, and hence I felt that it was a propitious time to issue a new consistent set of values of the general constants. During the writing of the paper further experimental progress has been made, and one can now (October 1941) say, with some assurance, that there is a good agreement between the direct and indirect values of h/e.

Even two years ago I felt that the preponderance of evidence was in favour of the indirect value, and hence in this paper, as in 1939,* I merely *assume* that the experimental values of e and e/m are correct, and that the Rydberg constant formula is valid. Then h/e (and h) becomes a *derived* constant. A brief comparison of this indirectly calculated value of h/e with the latest experimental values is given in Section M.

From the wave-lengths of x-ray lines, measured in I.A. units with a ruled grating, and the observed angles θ at which these same x-ray lines are reflected from a crystal, one can calculate by Bragg's equation ($\lambda = 2d \sin\theta$) the value of d, the effective grating space of the crystal. The calculation is best made by comparing the observed grating wave-lengths of x-ray lines with their observed wave-lengths on the Siegbahn system. In this system it is *assumed* that the effective grating space of calcite at 18° c., for first-order reflection, is $d''_{18} = 3{\cdot}02904 \times 10^{-8}$ cm. After correction for refraction † one obtains for the true grating space, $d'_{18} = 3{\cdot}029449 \times 10^{-8}$ cm., at 18° c., and $d'_{20} = 3{\cdot}029512 \times 10^{-8}$ cm., at 20° c. The probable errors are negligibly small.

These values are all on the Siegbahn system. To convert them to the ruled grating (or I.A.) system, one multiplies by $1{\cdot}002034 \pm 0{\cdot}000060$. This is the value calculated by Bearden (1935; see also Bearden, 1941) for the ratio λ_g/λ_s of x-ray wave-lengths on the grating and Siegbahn systems. It represents the weighted average of all known determinations. Bearden obtains $\pm 0{\cdot}000016$ as

* In August 1939 a mimeographed table of constants, with a brief explanation, was circulated privately for suggestions and corrections, preparatory to a detailed published paper.

† See G. C. 1929, p. 40, for details.

the calculated probable error, based on external consistency.* To allow for systematic errors of all kinds, I have assumed a probable error nearly four times as large. We thus obtain $d_{20} = (3{\cdot}03567_4 \pm 0{\cdot}00018) \times 10^{-8}$ cm. for the true grating space, at 20° C., on the ruled grating system—which we identify with the c.g.s. system.†

Knowing the value of d_{20}, we can now calculate the Avogadro number N_0 from the equation

$$N_0 = M/\{2\rho\phi(d_{20})^3\}, \qquad \ldots\ldots(1)$$

where M = molecular weight of calcite ($CaCO_3$), ρ = density of calcite at 20° C., and ϕ = geometrical constant for calcite.‡ To get equation (1) we must make the crucial assumption that the calcite crystal has a geometrically perfect structure. Because of the original apparent discrepancy between the oil-drop and the grating values of e, this assumption regarding the calcite crystal has been subjected to prolonged experimental and theoretical investigation. The beautiful work of DuMond and Bollman (1936, 38) now seems to have decided the assumption in the affirmative, although *very small* deviations from a geometrically perfect structure are still not excluded.§

The numerical values to be used in equation (1) are as follows. From the atomic weights listed in table *a'* we obtain $M = 100{\cdot}091_4 \pm 0{\cdot}005$ (chemical scale). With essentially no discussion of the evidence, I adopt $\rho = 2{\cdot}71029 \pm 0{\cdot}00003$ g.cm.$^{-3}$ at 20° c., and $\phi = 1{\cdot}09594 \pm 0{\cdot}00001$.‖ With these values we get, as the *adopted* value of the Avogadro number on the chemical scale,

$$N_0 = (6{\cdot}0228_3 \pm 0{\cdot}0011) \times 10^{23} \text{ mole}^{-1}.$$

With our adopted value $F = 9648{\cdot}7_7 \pm 1{\cdot}0$ abs. e.m.u. (Section J) we derive ¶

* A very recent summary of all data by Bearden (1941) leads to $\lambda_0/\lambda_4 = 1{\cdot}00203 \pm 0{\cdot}00001$, which does not differ essentially from his 1935 average of $1{\cdot}002034 \pm 0{\cdot}000016$.

† More precisely, the result is 3·03567 I.A. units, but for the present purpose any difference between an I.A. unit and 10^{-8} cm. is completely negligible.

‡ See Bearden (1931) for the equations that connect ϕ with the cleavage angles of calcite.

§ It is quite possible that, in order to allow for such a small mosaic effect, the probable errors adopted here for N_0 and e should have been taken twice as large. See Miller and DuMond (1940)

‖ Bearden (1931) adopts 2·71030 as the best average of the latest precision results. But his recalculated value (2·71034) of the De Foe and Compton data should be 2·71030, and this result is then to be averaged with Bearden's own result, 2·71026, and a later result, 2·71029, obtained with the same crystals at the National Bureau of Standards (see Bearden, 1938). Hence our adopted average and its probable error seem justified. The value of ϕ and its probable error, which corresponds to 2″ of arc in the angle β, come directly from Bearden (1931).

¶ A radical, but apparently necessary, revision of symbols is being made here. In the table of derived constants, it is essential that each symbol represent not only a specified quantity (such as the charge on the electron), but also a specified quantity in terms of a *specified unit*. In G.C. 1929 and elsewhere I have followed the objectionable but common practice of using e for the charge in e.s.u. and e/m for the specific charge in e.m.u. Since almost all precision determinations of electric and magnetic quantities are made directly in practical units, which can be converted to e.m.u. by a mere shift of decimal point, I propose now to use *unprimed* symbols such as e and F to represent these quantities in e.m.u., and the *primed* symbols e' and F' to represent them in e.s.u. Then $e' = ec$, $F' = Fc$, $e'/m = ec/m$, $h/e' = h/(ec)$, etc. F and e/m thus have the same meaning as in G.C. 1929, but e' corresponds to the previous e, and h/e' to the previous h/e.

$$e = F/N_0 = (1{\cdot}60203_3 \pm 0{\cdot}00034) \times 10^{-20} \text{ abs. e.m.u.}$$

and

$$e' = ec = (4{\cdot}8025_1 \pm 0{\cdot}0010) \times 10^{-10} \text{ abs. e.s.u.}$$

The value of e, or of e', when derived from oil-drop work, represents a *fundamental* constant, and N_0 is then a *derived* constant, according to the equation $N_0 = F/e$. The two constants N_0 and e are so classified in G.C. 1929 and in practically all more recent discussions. But when e is derived in the way now discussed, it is not e but N_0 that constitutes the fundamental constant, and $e\,(=F/N_0)$ then becomes a *derived* constant. Hence e should logically now appear in table *c*, along with other derived constants. It seems advisable, however, to retain it in table *a*.*

The " grating " value of e, as just calculated, is certainly more accurate than any value that can now be obtained from oil-drop work. This statement remains true, even if very much larger probable errors are assumed for the various constants of calcite, as Kirchner (1939) has done. As already stated, no discussion will be included here of recent oil-drop work, or of work on the viscosity of air, but it may be noted that a summarized table of such work has been given by Robinson (1938). Subsequent to his paper, Bearden (1939 b) has obtained $1834{\cdot}16 \times 10^{-7}$ c.g.s. units for the 23° C. viscosity, which with Millikan's oil-drop data yields $e' = 4{\cdot}8129 \times 10^{-10}$ abs. e.s.u., as compared with 4·768 given by the viscosity value used by Millikan. Apart from any error in the viscosity measurement, the probable error in e is at least $\pm 0{\cdot}005 \times 10^{-10}$ e.s.u., due to errors in Millikan's oil-drop work.

Hopper and Laby (1941) have just published the results of a new oil-drop measurement of e. Their published result corresponds to a value of e higher by $0{\cdot}0054 \times 10^{-10}$ e.s.u. than that given by Millikan's data. They conclude that in their work the error in e, exclusive of any error in the viscosity of air, is only $\pm 0{\cdot}0013 \times 10^{-10}$ e.s.u. But they also conclude that the present best value of the viscosity of air is $(1830{\cdot}0 \pm 2{\cdot}5) \times 10^{-7}$ c.g.s. units—and they thus get $e = (4{\cdot}802 \pm 0{\cdot}010) \times 10^{-10}$ abs. e.s.u. for the final result of their work. Although their final value of e thus agrees almost exactly with our adopted x-ray value, the probable error is ten times as large, and, moreover, this large error is due almost entirely, according to Hopper and Laby, to the error in the viscosity of air, rather than to errors in their own oil-drop measurements.

SECTION L

THE SPECIFIC CHARGE OF THE ELECTRON (e/m)

The value of e/m has been measured in numerous investigations, and by the use of many different methods. In fact no other atomic constant has been measured so often and in so many different ways. The entire situation regarding

* The foregoing considerations do not affect the calculated value of N_0 (or of e), but they do affect the calculated probable error of N_0. Thus if one should calculate the probable error in N_0 from the stated probable errors in F and e, on the assumption that $N_0 = F/e$, the probable error in F would thereby enter twice, whereas by equation (1) it does not enter at all.

e/m has, however, changed since G.C. 1929 was written. At that time it appeared to me necessary to adopt *two* values of e/m, a " deflection " value of $(1{\cdot}769 \pm 0{\cdot}002) \times 10^7$ abs. e.m.u. and a " spectroscopic " value of $1{\cdot}761 \pm 0{\cdot}001$. The apparent necessity for this unusual procedure is illustrated by the fact that Gerlach (1926), in a complete and authoritative review, adopted 1·766 as the best average, and listed in his final summary *no* value under 1·763. Possibly as a result of the inconsistency emphasized in G.C. 1929, there have since appeared numerous investigations, both experimental and theoretical, and from time to time I have reported (Birge, 1932 c, 34 a, 36, 38, 40 b) on the resulting situation. At the present time it is quite clear that the " low " spectroscopic value of G.C. 1929 is much more nearly correct than the higher value advocated by Gerlach and others.

The most important spectroscopic methods for evaluating e/m involve the Rydberg constant (R). A detailed paper on the best values of R and of e/m, from the spectra of H, D and He^+, has now been prepared (Birge, 1941), and the reader is referred to that paper for the relevant material. In connection with its preparation I have recalculated much more thoroughly than at any time in the past all of the precision spectroscopic work bearing on the subject. As a result o fan important investigation by Drinkwater, Richardson and Williams (1940), there has been a significant change in the value of R_∞ from that adopted in G.C. 1929.* The recent change in the value of q (Section E) also produces a small change in *every* value of e/m.

At the present time there are available twelve precision values of e/m, obtained by seven essentially different methods. Table 2 is reproduced from the paper (Birge, 1941) just mentioned. In the case of each result, the published data have been recalculated with the values of the auxiliary constants adopted

Table 2. Values of e/m

	e/m	Probable error	Author	Method	$f(F_{iq})$
1	1·75913	$3{\cdot}7 \times 10^{-4}$	Drinkwater, Richardson and W. E. Williams (1940)	H_α and D_α	F_{4q}
2	1·75797	$5{\cdot}0 \times 10^{-4}$	R. C. Williams (1938)	H_α and D_α	F_{4q}
3	1·75914	$5{\cdot}0 \times 10^{-4}$	C. F. Robinson (1939)	H_α and D_α	F_{4q}
4	1·75815	$6{\cdot}0 \times 10^{-4}$	Shane and Spedding (1935)	H_α and D_α	F_{4q}
5	1·76048	$5{\cdot}8 \times 10^{-4}$	Houston (1927); Chu (1939)	R^B and R_{He}	F_{4q}
6	1·75700	$7{\cdot}0 \times 10^{-4}$	Kinsler and Houston (1934)	Zeeman effect	$1/q$
7	1·76006	$4{\cdot}0 \times 10^{-4}$	Bearden (1938)	X-ray refraction	$1/F_{4q}$
8	1·76110	$10{\cdot}0 \times 10^{-4}$	Perry and Chaffee (1930)	Direct velocity	$1/pq$
9	1·75900	$9{\cdot}0 \times 10^{-4}$	Kirchner (1931, 32)	Direct velocity	$1/pq$
10	1·75982	$4{\cdot}0 \times 10^{-4}$	Dunnington (1933, 37)	Magnetic deflection	$1/q$
11	1·75820	$13{\cdot}0 \times 10^{-4}$	Shaw (1938)	Crossed fields	p/q
12	1·75870	$8{\cdot}0 \times 10^{-4}$	Goedieke (1939)	Busch method	p/q

* The various values of R now adopted (all on the I.A. scale) are, $R_H = 109677{\cdot}581_8 \pm 0{\cdot}007_8$ cm.$^{-1}$, $R_D = 109707{\cdot}419_3 \pm 0{\cdot}007_8$ cm.$^{-1}$, $R_{He} = 109722{\cdot}263 \pm 0{\cdot}012$ cm.$^{-1}$, and finally, $R_\infty = 109737{\cdot}303 \pm 0{\cdot}017$ cm.$^{-1}$ (or $\pm 0{\cdot}05$ cm.$^{-1}$ in the c.g.s. system).

here. In the case of the spectroscopic results, several modifications have been made in the detailed reduction of the data. The final column of the table lists the dependence of the value of e/m on F_i, the value of the Faraday in international units (hence $F_i q$ gives the Faraday in absolute units), and on p and q (of Section E).

Several significant conclusions may be drawn from the table. In the first place, the weighted average of all twelve values of e/m is $1{\cdot}7592_{98} \pm 0{\cdot}0001_{92}$ ($\pm 0{\cdot}0001_{58}$).* The ratio r_e/r_i is 1·213, which is quite satisfactory. In other words, there is no longer any serious discrepancy between the various results, considered collectively. There is, however, still a very slight discrepancy between the "spectroscopic" and the "deflection" results. Thus the weighted average of the first six † results listed in table 2 (the "spectroscopic" values) is $e/m = 1{\cdot}75880 \pm 0{\cdot}00028_2$ ($\pm 0{\cdot}00020_8$), and $r_e/r_i = 1{\cdot}36$. The last five results of the table (the "deflection" values) give an average of $1{\cdot}75959 \pm 0{\cdot}00024_1$ ($\pm 0{\cdot}00030_6$), and $r_e/r_i = 0{\cdot}79$. Hence each set of results is reasonably self-consistent. The discrepancy of 0·00079 between the two averages, although only one-tenth of the apparent discrepancy in 1929, is still 2·1 times its own present probable error, whether the error is calculated from internal or from external consistency. Since the odds are about one in six that such a discrepancy is due to pure chance, the situation is not entirely satisfactory. In view, however, of the many serious sources of systematic error that have been revealed, from time to time, in almost every method for measuring e/m, I think that the small remaining discrepancy should not be taken too seriously.

The probable error of the weighted average of all twelve results is only $\pm 0{\cdot}0002$, as already stated, but just because of the discrepancy of 0·0008 between the two types of measurement, I have increased the adopted probable error to $\pm 0{\cdot}0005$. Hence the final *adopted* result is

$$e/m = (1{\cdot}7592 \pm 0{\cdot}0005) \times 10^7 \text{ abs.-e.m.u. g}^{-1}$$

$$e'/m = ec/m = (5{\cdot}2736_6 \pm 0{\cdot}0015) \times 10^{17} \text{ abs.-e.s.u. g}^{-1}$$

The data listed in the table make possible an independent evaluation of the Faraday, since the first five spectroscopic results involve F and the five deflection results do not. The new spectroscopic average is 1·75897, and the result is, therefore,

$$F = (9651{\cdot}40)(1{\cdot}75959)/(1{\cdot}75897) = 9654{\cdot}80 \text{ abs. e.m.u. (physical scale)},$$

or 96535 int. coulombs (chemical scale).‡ Purely on the basis of the external consistency of the various e/m values involved in the two averages, the resulting

* Even as early as 1932 I suggested (Birge, 1932 b) 1·7592 as the best value of e/m. But in 1934 this "constant" had reached an "all-time low" of 1·757 (Birge, 1934 a). No one, however, has as yet suggested that the value of e/m is actually changing with time, contrary to the situation in the case of the velocity of light (Section A)!

† Bearden (1939 a) objects to my classification (Birge, 1938) of his x-ray refraction value (Bearden (1938), table 2) with the spectroscopic results. Hence I now place it in a separate category of its own.

‡ The corresponding corrected value of e/m is then necessarily just the deflection average 1·75959, since F is not involved in that average.

value of F has a probable error of ± 20 int. coulombs. In Section J the best average electrochemical value of F has been found to be $96501{\cdot}2 \pm 3{\cdot}0$ int. coulombs, although the adopted probable error has been taken as ± 10 int. coulombs, to allow for possible unforeseen errors. It thus seems fairly certain that the new method for evaluating F is not as accurate as the usual method, and, furthermore, it is far more likely, in my opinion, that the adopted value of 96501 int. coulombs is *too high*, rather than too low.

As noted in Section J, it is now proposed to use the adopted values of e and e/m in the Rydberg constant formula, to *calculate* a value of h. The Birge-Bond diagram immediately shows* that the adopted value of e/m, even with a probable error as large as $\pm 0{\cdot}0005$, is by far the most accurately determined function involving the atomic constants. Hence it is the values of e and h/e that now require critical scrutiny, rather than that of e/m.

SECTION M

THE PLANCK CONSTANT h

As already stated in Section K, the Planck constant h, and also h/e, will be treated here as derived constants, and will be calculated from the measured values of e and e/m and the Bohr formula for the Rydberg constant R_∞. This calculation will be followed by a very brief discussion of recent experimental work on h/e.

The Rydberg constant formula involves e, m and h, but e, as calculated in Section K, is itself a derived constant, equal to F/N_0. Hence R_∞ is first to be expressed as a function of h, e/m, F and N_0, and then solved for h, or for h/e. The formula is normally written in the form

$$R_\infty = 2\pi^2(e')^5/h^3c(e'/m) \text{ cm}^{-1}, \qquad \ldots\ldots(1)$$

where e' indicates the electronic charge in e.s.u. We now change to e.m.u., substitute F/N_0 for e, and in order to simplify the equation introduce the symbol

$$B = 2\pi^2c^3/R_\infty. \qquad \ldots\ldots(2)$$

Then equation (1) may be written

$$h = \left\{\frac{BF^5}{N_0^5(e/m)}\right\}^{1/3} \qquad \ldots\ldots(3)$$

or

$$h/e = \left\{\frac{BF^2}{N_0^2(e/m)}\right\}^{1/3} \qquad \ldots\ldots(4)$$

Our adopted values of the necessary constants are

$$R_\infty = 109737{\cdot}30_3 \pm 0{\cdot}05 \text{ cm}^{-1}$$
$$c = (2{\cdot}99776 \pm 0{\cdot}00004) \times 10^{10} \text{ cm. sec.}^{-1}$$
$$F = 9648{\cdot}7_7 \pm 1{\cdot}0 \text{ abs.-e.m u. g.-equiv.}^{-1}$$
$$N_0 = (6{\cdot}0228_3 \pm 0{\cdot}0011) \times 10^{23} \text{ mole.}^{-1}$$
$$e/m = (1{\cdot}7592 \pm 0{\cdot}0005) \times 10^7 \text{ abs.-e.m.u. g.}^{-1}$$

* See, for instance, Dunnington (1939).

Hence, from equation (2),

$$B = (4{\cdot}84580_{617} \pm 0{\cdot}00019_4) \times 10^{27} \text{ cm}^4 \text{ sec}^{-3}$$

and from equations (3) and (4),

$$h = (6{\cdot}624_3 \pm 0{\cdot}002_4) \times 10^{-27} \text{ erg. sec.}$$

$$h/e = (4{\cdot}1349_0 \pm 0{\cdot}0007_1) \times 10^{-7} \text{ erg. sec. abs.-e.m.u.}^{-1}$$

$$h/e' = (1{\cdot}3793_3 \pm 0{\cdot}0002_3) \times 10^{-17} \text{ erg. sec. abs.-e.s.u.}^{-1}$$

As mentioned in section K, there are a number of functions involving e, m and h that have been measured with considerable precision. The reader is referred to the papers already quoted (Wensel, 1939; Dunnington, 1939; DuMond, 1939, 40; Kirchner, 1939) for the necessary details. Although I have rather recently carried out extensive and as yet unpublished calculations on the atomic constants, I must here content myself with the following very brief remarks.

The indirect calculation just made corresponds, on the Birge-Bond diagram, to passing a straight line through the two points: (1) the observed value of e, which is plotted at abscissa $p = 0$, and (2) the observed value of e/m, which—after reduction to a $f(e, h)$ only, by means of the Rydberg constant formula—is plotted at $p = 3/5$. Then the corresponding value of h/e lies on the straight line, at $p = 1{\cdot}0$.

Instead of using only these two points, in order to calculate an indirect value of h/e, we may use a group of points and pass, by least squares, the best straight line through them. All such possible points are listed and discussed by Dunnington (1939), but the only ones, in addition to e and e/m, that now seem to be sufficiently accurate to justify real consideration in such a least squares solution are the following *: (1) von Friesen (1935) has obtained a value of $(e/m)(h/e)^2$, from electron diffraction experiments. This point is plotted at $p = 1/3$. (2) Christy and Keller (1940; see also Birge, 1940) have obtained a value of the fine-structure constant α from all available x-ray data, as critically compiled by the writer (Birge, 1939 b). This point is plotted at $p = 1/2$. (3) The value of h/m has been measured by Ross and Kirkpatrick (1934), from the Compton shift. When m is eliminated by means of the Rydberg constant equation, one gets a $f(e, h)$ that is also to be plotted at $p = 1/2$. Hence these two observed points at $p = 1/2$ are first to be replaced by their weighted average, when a least squares solution is made. (4) H. R. Robinson and co-workers have evaluated $(e/m)(e/h)$ by means of the observed $H\rho$ values of electrons ejected by x rays. The two latest investigations, which give almost identical results, are those by Mayo and Robinson (1939) and by Clews and Robinson (1940). This point is plotted at $p = 2/3$.

We thus have points at $p = 0$, $1/3$, $1/2$, $3/5$ and $2/3$. With the values of e and e/m adopted in August 1939, which are almost identical with the present adopted values, and with the use of the additional points just mentioned, I have

* See figure 1 for these points.

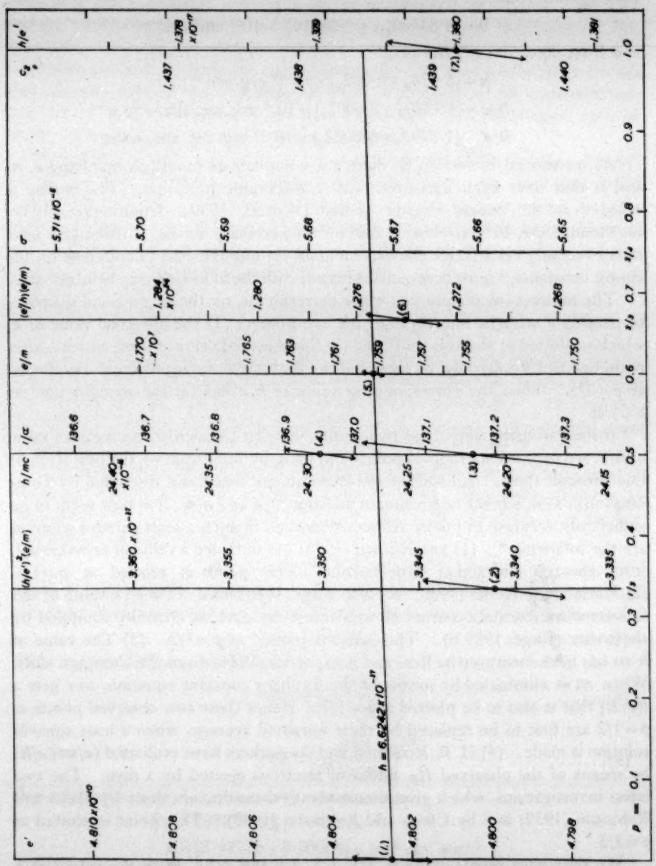


Figure 1. Nomographic chart (Birge-Bond diagram) showing relations between various $f(e, m, h)$. The ordinate scale is so highly magnified (only $\frac{1}{2}$% total variation in e') that many less accurately determined $f(e, m, h)$ fall outside the limits shown. The values of the constants adopted in this paper are given by the straight line through points (1) and (5), whose slope corresponds to $h = 6{\cdot}6242 \times 10^{-27}$. The diagram is based on $h = 6{\cdot}625$ for a horizontal line.

obtained for the least-squares solution a straight line that almost coincides with that passing exactly through the e and e/m points at $p = 0$ and 3/5. Furthermore, no point deviates from this line by an amount appreciably greater than its apparent probable error. Hence all of the present most accurate data involving functions of e, m and h—with the exception of the direct measurements of h/e—are very consistent, and they collectively yield a value of h/e almost identical with the value just adopted.

When, however, we consider the various direct measurements of h/e the picture is less satisfactory. One can determine h/e from (1) photoelectric effect, (2) ionization and excitation potentials of atoms, (3) the radiation constant $c_2 = (h/e)(Fc/R_0)$, and (4) the Duane-Hunt limit of the continuous x-ray spectrum. A new and interesting method, involving a comparison of the heat of dissociation of iodine measured in cm^{-1} units per atom with that measured in calories per mole, has recently been used by Perlman and Rollefson (1941).

Of these five methods for determining h/e, only the fourth appears to have attained the accuracy necessary for our purpose. Until very recently the most trustworthy work using this method was apparently that by DuMond and Bollman (1937). Their result, in terms of the present values of the auxiliary constants, is $h/e' = (1{\cdot}3765 \pm 0{\cdot}0003) \times 10^{-17}$, and the divergence between this result and our indirect value of $(1{\cdot}3793 \pm 0{\cdot}0002_3) \times 10^{-17}$ constitutes the fundamental discrepancy already mentioned more than once in this paper. The discrepancy is discussed fully by DuMond and Bollman (1937), in terms of the Birge-Bond diagram.

More recently Ohlin (1941) has found * $h/e' \pm (1{\cdot}3785 \pm 0{\cdot}0005) \times 10^{-17}$. By lowering the pressure in his x-ray tube, he eliminates entirely the troublesome " tail " of the excitation curve, found by all previous investigators. On the other hand, he fails to correct his voltage for the work function of tungsten (about 4·5 volts). The discussion of DuMond and Bollman (1937) certainly indicates the necessity for this correction, and, if it is made, Ohlin's new average result

* Ohlin obtains his stated uncertainty of $\pm 0{\cdot}0008$ by *adding* his various estimated errors instead of by taking the square root of the sum of the squares. The use of my auxiliary constants changes his published 1·3787 to 1·3785.

The experimental points shown (each with its adopted probable error indicated by a double arrow) are :

(1) $e' = (4{\cdot}8025_1 \pm 0{\cdot}0010) \times 10^{-10}$ e.s.u. (adopted value).
(2) von Friesen (1935), $(e/m)(h/e')^2 = (3{\cdot}3410_8 \pm 0{\cdot}0037) \times 10^{-27}$ in " mixed units ".
(3) Ross and Kirkpatrick (1934), $h/mc = (24{\cdot}216 \pm 0{\cdot}057) \times 10^{-11}$ cm.
(4) Christy and Keller (1940) and Birge (1939 b), $1/\alpha = 136{\cdot}95 \pm 0{\cdot}05$.
(5) $e/m = (1{\cdot}7592 \pm 0{\cdot}0005) \times 10^7$ e.m.u. (adopted value).
(6) Mayo and Robinson (1939) and Clews and Robinson (1940), $(e/m)(e'/h) = (1{\cdot}2741 \pm 0{\cdot}0015) \times 10^{24}$ in " mixed " units.
(7) Ohlin (1941), DuMond, Bailey, Panofsky and Green (1941), and DuMond, Panofsky and Green (1941), $h/e' = (1{\cdot}3800 \pm 0{\cdot}0005) \times 10^{-17}$ e.s.u.

The scales for the radiation constants σ and c_2 are also shown, at $p = \frac{1}{4}$ and 1·0, respectively The scale for $\mu_1 N_0$ (magnetic moment per mole for one Bohr magneton per molecule—see table c) should be plotted at $p = \frac{1}{2}$, and may be obtained by multiplying the Compton shift (h/mc) scale by $F'/4\pi = 2{\cdot}301754 \times 10^{13}$. The scales for many other $f(e, m, h)$ may similarly be plotted on such a diagram.

becomes 1·3800. Both the corrected and the uncorrected results are thus in good agreement with our indirect value.

DuMond and co-workers* are now engaged in remeasuring h/e. By the use of P. A. Ross balanced filters, in conjunction with a two-crystal monochromator, they also have completely eliminated the " tail " on the excitation curve, and they have obtained a remarkably linear plot from which the intercept can be read with very great accuracy. These investigators have now checked their voltage calibration with Bureau of Standards apparatus and their present (October 1941) experimental result † is $(1{\cdot}3800 \pm 0{\cdot}0008) \times 10^{-17}$, in exact coincidence with our corrected Ohlin result.

Bearden and Schwarz (1941) have also recently made a preliminary report on a measurement of h/e'. Their results vary from 1·3772 to 1·3793, according to the metal used. They, however, have not eliminated the tail on the excitation curve, and hence their results cannot be considered as trustworthy as those of Ohlin or of DuMond.

Although no discussion is given here of the values of h/e obtained by the remaining four, and less accurate, methods already listed, it is only fair to say that *all* such investigations, with the exception of that by Perlman and Rollefson (1941), have thus far yielded values of h/e' that are considerably *below* the indirect value of 1·3793, and in some cases considerably below even the older DuMond and Bollman (1937) value of 1·3765.‡ On the other hand, the very recent work on h/e just discussed has brought to light important sources of possible systematic error in the earlier work, and furthermore, the new work, as we have seen, is giving values of h/e as high as, or even higher than, our " high " indirect value. Hence it is possible to assume that no real conflict exists between the direct and the indirect values of h/e, and if this assumption is finally confirmed, the values of the general physical constants will be at last in a fairly satisfactory condition.

With these all too brief remarks, I close the present discussion of the general constants. Tables *a*, *a'* and *b*, which follow, are reasonably complete, but table *c* represents a selection of a few only of the derived constants in most common use. A far more extensive list will be published in connection with the detailed article that I hope to complete in the not too distant future, provided the international situation allows any pure scientific work to be continued.

Table *a*. Principal constants and ratios §

Section		
A	Velocity of light	$c = (2{\cdot}99776 \pm 0{\cdot}00004) \times 10^{10}$ cm.sec.$^{-1}$
B	Gravitation constant	$G = (6{\cdot}670 \pm 0{\cdot}005) \times 10^{-8}$ dyne.cm. g.$^{-2}$
C	Litre (=1000 ml.)	$l = 1000{\cdot}028 \pm 0{\cdot}002$ cm.^{3}

* DuMond, Bailey, Panofsky and Green, 1941 ; DuMond, Panofsky and Green, 1941.

† I am greatly indebted to Dr. DuMond and Mr. Panofsky for prompt and detailed reports on the progress of this work.

‡ See DuMond (1940, p. 463) for an extended table of such results.

§ Unless otherwise specified, all quantities in these tables that involve the mole or the gram equivalent are on the chemical scale of atomic weights.

Section		
D	Volume of ideal gas (0° c., A_0)	$V_0=(22{\cdot}4146\pm0{\cdot}0006)\times10^3$ cm.^{3}atmos. mole^{-1}
		$V_0'=22{\cdot}4140\pm0{\cdot}0006$ litre.atmos.mole^{-1}
D	Volume of ideal gas (0° c., A_{45})	$V_{45}=(22{\cdot}4157\pm0{\cdot}0006)\times10^3$ cm.^{3} atmos. mole^{-1}
		$V_{45}'=22{\cdot}4151\pm0{\cdot}0006$ litre.atmos.mole^{-1}
E	International ohm (= p abs. ohm)	$p=1{\cdot}00048\pm0{\cdot}00002$
E	International ampere (= q abs. amp.)	$q=0{\cdot}99986\pm0{\cdot}00002$
F	Atomic weights (see table *a'*)	
G	Standard atmosphere	$A_0=(1{\cdot}013246\pm0{\cdot}000004)\times10^6$ dyne. cm.$^{-2}$atmos.$^{-1}$
G	45° atmosphere	$A_{45}=(1{\cdot}013195\pm0{\cdot}000004)\times10^6$ dyne. cm.$^{-2}$atmos.$^{-1}$
H	Ice-point (absolute scale)	$T_0=273{\cdot}16\pm0{\cdot}01°$ K.
I	Joule equivalent	$J_{15}=4{\cdot}1855\pm0{\cdot}0004$ abs.-joule.cal.$_{15}^{-1}$
I	Joule equivalent (electrical)	$J_{15}'=4{\cdot}1847\pm0{\cdot}0003$ int.-joule.cal.$_{15}^{-1}$
J	Faraday constant	
	(1) Chemical scale :	
		$F=96501{\cdot}_2\pm10$ int.-coul. g.-equiv.$^{-1}$
		$=96487{\cdot}_7\pm10$ abs.-coul. g.-equiv.$^{-1}$
		$=9648{\cdot}7_2\pm1{\cdot}0$ abs.-e.m.u. g.-equiv.$^{-1}$
		$F'=Fc=(2{\cdot}89247\pm0{\cdot}00030)\times10^{14}$ abs.-e.s.u. g.-equiv.$^{-1}$
	(2) Physical scale :	
		$F=96514{\cdot}_0\pm10$ abs.-coul. g.-equiv.$^{-1}$
		$=9651{\cdot}4_0\pm1{\cdot}0$ abs.-e.m.u. g.-equiv.$^{-1}$
		$F'=Fc=(2{\cdot}89326\pm0{\cdot}00030)=10^{14}$ abs.-e.s.u. g.-equiv.$^{-1}$
K	Avogadro number (chemical scale)	$N_0=(6{\cdot}0228_3\pm0{\cdot}0011)\times10^{23}$ mole^{-1}
K	Electronic charge	$e=F/N_0=(1{\cdot}60203_3\pm0{\cdot}00034)\times10^{-19}$ abs. e.m.u.
		$e'=ec=(4{\cdot}8025_1\pm0{\cdot}0010)\times10^{-10}$ abs. e.s.u.
L	Specific electronic charge	$e/m=(1{\cdot}7592\pm0{\cdot}0005)\times10^7$ abs.-e.m.u. g.$^{-1}$
		$e'/m=ec/m=(5{\cdot}2736_4\pm0{\cdot}0015)\times10^{17}$ abs.-e.s.u. g.$^{-1}$
M	Planck constant	h (see table *c*).

Table *a'*. Atomic weights

(1) Physical scale ($^{16}O=16{\cdot}0000$)

$^1H=1{\cdot}00813\pm0{\cdot}00001_7$ $^2H=2{\cdot}01473\pm0{\cdot}00001_9$

$H=1{\cdot}00827_6\pm0{\cdot}00001_7$ (from $^1H/^2H$ abundance 6900 ± 100)

$^4He=4{\cdot}00389\pm0{\cdot}00007$

$^{12}C=12{\cdot}00386\pm0{\cdot}00004$ $^{13}C=13{\cdot}00761\pm0{\cdot}00015$

$C=12{\cdot}01465\pm0{\cdot}00023$ (from $^{12}C/^{13}C$ abundance $=92\pm2$)

$^{14}N=14{\cdot}00753\pm0{\cdot}00005$ $^{15}N=15{\cdot}0049\pm0{\cdot}0002$

$N=14{\cdot}01121\pm0{\cdot}00009_3$ (from $^{14}N/^{15}N$ abundance $=270\pm6$)

$^{16}O=16{\cdot}0000$ $^{17}O=17{\cdot}0045$ $^{18}O=18{\cdot}0049$

$O=16{\cdot}00435_7\pm0{\cdot}00008_8$

(from abundance $^{16}O:{}^{18}O:{}^{17}O=(506\pm10):1:(0{\cdot}204\pm0{\cdot}008)$)

(2) Chemical scale (O = 16·0000)

Ratio physical to chemical scale

$r = (16{\cdot}004357 \pm 0{\cdot}000086)/16 = 1{\cdot}000272 \pm 0{\cdot}000005$

$^1H = 1{\cdot}00785_8 \pm 0{\cdot}00001_8$ (from physical scale)
$^2H = 2{\cdot}01418_2 \pm 0{\cdot}00002_1$ (from physical scale)
$H = 1{\cdot}00800_2 \pm 0{\cdot}00001_8$ (from physical scale)
$^4He = 4{\cdot}00280 \pm 0{\cdot}00007$ (from physical scale)
$C = 12{\cdot}01139 \pm 0{\cdot}00024$ (from physical scale)
$N = 14{\cdot}00740 \pm 0{\cdot}00012$ (from physical scale)
$N = 14{\cdot}0086 \pm 0{\cdot}0007$ (direct observation)
$Na = 22{\cdot}994 \pm 0{\cdot}003$
$Cl = 35{\cdot}457 \pm 0{\cdot}001$
$Ca = 40{\cdot}080 \pm 0{\cdot}005$
$Ag = 107{\cdot}880 \pm 0{\cdot}002$
$I = 126{\cdot}915 \pm 0{\cdot}004$

Table *b*. Additional quantities evaluated or used in connection with table *a*

Section		
A	Ratio of e.s.u. to e.m.u. (direct)	$c' = (2{\cdot}9971_9 \pm 0{\cdot}0001) \times 10^{10}$ cm.$^{1/2}$sec.$^{-1/2}$ int.-ohm$^{1/2}$ $= (2{\cdot}9978_4 \pm 0{\cdot}0001_6) \times 10^{10}$ cm.sec.$^{-1}$
A	Ratio of e.s.u. to e.m.u. (indirect)	$c' = c = (2{\cdot}99776 \pm 0{\cdot}00004) \times 10^{10}$ cm.sec.$^{-1}$
B	Average density of earth	$\delta = 5{\cdot}517 \pm 0{\cdot}004$ g.cm.$^{-3}$
C	Maximum density of water	$\delta_m(H_2O) = 0{\cdot}999972 \pm 0{\cdot}000002$ g.cm.$^{-3}$
D	Acceleration of gravity (standard)	$g_0 = 980{\cdot}665$ cm.sec.$^{-2}$
D	Acceleration of gravity (45°)	$g_{45} = 980{\cdot}616$ cm.sec.$^{-2}$
D	Density of oxygen gas (0° c., A_{45})	$L_1 = 1{\cdot}42897 \pm 0{\cdot}00003$ g. litre^{-1}
D	Limiting density of oxygen gas (0° c., A_{45})	$L_{\text{lim}} = 1{\cdot}427609 \pm 0{\cdot}000037$ g. litre^{-1}
D	Factor converting oxygen (0° c., A_{45}) to ideal gas	$1 - \alpha = 1{\cdot}000953_5 \pm 0{\cdot}000009_4$
E	International coulomb (= q abs. coul.)	$q = 0{\cdot}99986 \pm 0{\cdot}00002$
E	International gauss (= q abs. gauss)	
E	International henry (= p abs. henry)	$p = 1{\cdot}00048 \pm 0{\cdot}00002$
E	International volt (= pq abs. volt)	$pq = 1{\cdot}00034 \pm 0{\cdot}00003$
E	International joule (= pq^2 abs. joule)	$pq^2 = 1{\cdot}00020 \pm 0{\cdot}00004_5$
G	Specific gravity of Hg (0° c.; A_0) referred to air-free water at maximum density	$\rho_0 = 13{\cdot}59542 \pm 0{\cdot}00005$
G	Density of Hg (0° c., A_0)	$D_0 = 13{\cdot}59504_0 \pm 0{\cdot}00005_7$ g.cm.$^{-3}$
J	Electrochemical equivalents (chemical scale):	
	Silver (apparent)	$E^*_{Ag} = 1{\cdot}11800 \times 10^{-3}$ g. int.-coul.$^{-1}$
	(corrected)	$E_{Ag} = (1{\cdot}11807 \pm 0{\cdot}00012) \times 10^{-3}$ g.abs.-coul.$^{-1}$
	Iodine (apparent)	$E^*_I = (1{\cdot}315026 \pm 0{\cdot}000025) \times 10^{-3}$ g.int.-coul.$^{-1}$
	(corrected)	$E_I = (1{\cdot}31535 \pm 0{\cdot}00014) \times 10^{-3}$ g.abs.-coul.$^{-1}$

Section		
K	Effective calcite grating space (18° c.), Siegbahn system	$d''_{18} = 3{\cdot}02904 \times 10^{-8}$ cm.
K	True calcite grating space (20° c.), Siegbahn system	$d'_{20} = 3{\cdot}02951_8 \times 10^{-8}$ cm.
K	True calcite grating space (20° c.), c.g.s. system	$d_{20} = (3{\cdot}03567_4 \pm 0{\cdot}00018) \times 10^{-8}$ cm.
K	Ratio of grating and Siegbahn scales of wave-lengths	$\lambda_g/\lambda_s = 1{\cdot}002034 \pm 0{\cdot}000060$
K	Density of calcite (20° c.)	$\rho = 2{\cdot}71029 \pm 0{\cdot}00003$ g.cm.$^{-3}$
K	Structural constant of calcite (20° c.)	$\phi = 1{\cdot}09594 \pm 0{\cdot}00001$
K	Molecular weight of calcite (chemical scale)	$M = 100{\cdot}091_4 \pm 0{\cdot}005$
L	Rydberg constant for hydrogen (^{1}H)	$R_H = 109677{\cdot}581_3 \pm 0{\cdot}007_5$ cm.$^{-1}$ (I.A. scale)
L	Rydberg constant for deuterium (^{2}H)	$R_D = 109707{\cdot}419_3 \pm 0{\cdot}007_5$ cm.$^{-1}$ (I.A. scale)
L	Rydberg constant for helium	$R_{He} = 109722{\cdot}263 \pm 0{\cdot}012$ cm.$^{-1}$ (I.A. scale)
L	Rydberg constant for infinite mass	$R_\infty = 109737{\cdot}303 \pm 0{\cdot}017$ cm.$^{-1}$ (I.A. scale) or $\pm 0{\cdot}05$ cm.$^{-1}$ (c.g.s. system)

Table *c*. Partial list of derived quantities *

Planck constant :

$$h = \left\{\frac{2\pi^3 c^3 F^8}{R_\infty N_0^5 (e/m)}\right\}^{1/5} = (6{\cdot}624_3 \pm 0{\cdot}002_4) \times 10^{-27} \text{ erg. sec.}$$

$$h/e = \left\{\frac{2\pi^3 c^2 F^3}{R_\infty N_0^3 (e/m)}\right\}^{1/5} = (4{\cdot}1349_0 \pm 0{\cdot}0007_1) \times 10^{-7} \text{ erg. sec. abs.-e.m.u.}^{-1}$$

$$h/e' = h/ec = \left\{\frac{2\pi^3 F^3}{R_\infty N_0^3 (e/m)}\right\}^{1/5} = (1{\cdot}3793_3 \pm 0{\cdot}0002_3) \times 10^{-17} \text{ erg. sec. abs.-e.s.u.}^{-1}$$

Atomic weight of electron : $E = F/(e/m)$

(physical scale) $= (5{\cdot}4862_4 \pm 0{\cdot}0017) \times 10^{-4}$

(chemical scale) $= (5{\cdot}4847_5 \pm 0{\cdot}0017) \times 10^{-4}$

Band spectra constant connecting wave number and moment of inertia :

$$h/8\pi^2 c = \left\{\frac{F^4}{256\pi^4 R_\infty N_0^5 (e/m)}\right\}^{1/5} = (27{\cdot}98_{85} \pm 0{\cdot}01_0) \times 10^{-40} \text{g.cm.}$$

Boltzmann constant :

$$k = R_0/N_0 = V_0 A_0/T_0 N_0 = (1{\cdot}38047_4 \pm 0{\cdot}00026) \times 10^{-16} \text{erg. deg.}^{-1}$$

Charge in electrolysis of one gram of H :

$$F/\text{H} = 9572{\cdot}1_{73} \pm 1{\cdot}0 \text{ abs.-e.m.u. g}^{-1}$$

Charge in electrolysis of one gram of ^{1}H :

$$e/M_{1H} = F/^1\text{H} = 9573{\cdot}5_{80} \pm 1{\cdot}0 \text{ abs.-e.m.u. g.}^{-1}$$

* In order to be able to calculate, by propagation of errors, the probable error in a derived quantity, it is necessary to express the quantity explicitly in terms of the various fundamental quantities of table *a* or of table *b*, and that has been done in each case. Since in this paper *e* and *h* are treated as derived quantities, they do not therefore appear in such explicit expressions. But in calculating the *numerical* value of a derived quantity, the work can often be greatly simplified by using the values of other previously calculated derived quantities—in particular *e* and *h*. In order to show how certain derived quantities depend on quantities like *e* and *h*, such alternative expressions are given in many cases.

Compton shift at 90° :

$$h/mc = \left\{\frac{2\pi^2 F^2 (e/m)^2}{R_\infty N_0^2}\right\}^{1/3} = (0{\cdot}024265_{14} \pm 0{\cdot}000005_7) \times 10^{-8} \text{ cm.}$$

Energy in ergs of one abs.-volt-electron :

$$E_0 = 10^8 e = 10^8 F/N_0 = (1{\cdot}60203_3 \pm 0{\cdot}00034) \times 10^{-12} \text{ erg}$$

Energy in calories per mole for one abs.-volt-electron per molecule :

$$\frac{F(\text{abs. coul. per gram-equiv.})}{J_{15}\,(\text{abs. joules per cal.})} = 23052{\cdot}_{85} \pm 3{\cdot}_3 \text{ cal.}_{15}\text{. mole}^{-1}$$

Fine structure constant :

$$\alpha = 2\pi(e')^2/hc = \left\{\frac{4\pi R_\infty F(e/m)}{N_0}\right\}^{1/3} = (7{\cdot}2976_6 \pm 0{\cdot}0008_6) \times 10^{-3}$$

$$1/\alpha = 137{\cdot}030_9 \pm 0{\cdot}016$$

$$\alpha^2 = (5{\cdot}3256 \pm 0{\cdot}0013) \times 10^{-5}$$

Gas constant per mole :

$$R_0 = V_0 A_0/T_0 = (8{\cdot}31436 \pm 0{\cdot}00038) \times 10^7 \text{ erg. deg.}^{-1} \text{ mole}^{-1}$$

$$R_0' = R_0 \,.\, 10^{-7}/J_{15} = 1{\cdot}98646_7 \pm 0{\cdot}00021 \text{ cal.}_{15} \text{ deg.}^{-1} \text{ mole}^{-1}$$

$$R_0'' = V_0'/T_0 = (8{\cdot}20544_7 \pm 0{\cdot}00037) \times 10^{-2} \text{ litre. atmos. deg.}^{-1} \text{ mole}^{-1}$$

$$R_0''' = R_0/A_0 = V_0/T_0 = 82{\cdot}0566_9 \pm 0{\cdot}0037 \text{ cm.}^3\text{atmos. deg.}^{-1}\text{mole}^{-1};$$

also *

$$R_0 T_0 = V_0 A_0 = (2{\cdot}27115_0 \pm 0{\cdot}00006) \times 10^{10} \text{ erg. mole}^{-1}$$

Loschmidt number (0° C., A_0) :

$$n_0 = N_0/V_0 = (2{\cdot}6870_{12} \pm 0{\cdot}0005_0) \times 10^{19} \text{ atmos.}^{-1} \text{ cm.}^{-3}$$

Magnetic moment of one Bohr magneton :

$$\mu_1 = (h/4\pi)(e/m) = \frac{1}{4\pi}\left\{\frac{2\pi^2 c^3 F^3 (e/m)^2}{R_\infty N_0^3}\right\}^{1/3}$$

$$= (0{\cdot}9273_{48} \pm 0{\cdot}0003_7) \times 10^{-20} \text{ erg. gauss}^{-1}$$

Magnetic moment per mole for one Bohr magneton per molecule :

$$\mu_1 N_0 = \frac{1}{4\pi}\left\{\frac{2\pi^2 c^3 F^3 (e/m)^2}{R_\infty N_0^2}\right\}^{1/3} = 5585{\cdot}2_4 \pm 1{\cdot}6 \text{ erg. gauss.}^{-1} \text{ mole}^{-1}$$

Mass of α-particle :

$$M_\alpha = (\text{He} - 2\text{E})/N_0 = (6{\cdot}6442_3 \pm 0{\cdot}0012) \times 10^{-24} \text{ g.}$$

Mass of atom of unit atomic weight :

$$M_0 = 1/N_0 = (1{\cdot}66035 \pm 0{\cdot}00021) \times 10^{-24} \text{ g.}$$

Mass of electron :

$$m = e/(e/m) = (F/N_0)/(e/m) = (9{\cdot}1066_0 \pm 0{\cdot}0032) \times 10^{-28} \text{ g.}$$

Mass of ^{1}H atom :

$$M_{^1\text{H}} = {}^1\text{H}/N_0 = (1{\cdot}67339_3 \pm 0{\cdot}00031) \times 10^{-24} \text{ g.}$$

Mass of proton :

$$M_\text{P} = ({}^1\text{H} - \text{E})/N_0 = (1{\cdot}67248_3 \pm 0{\cdot}00031) \times 10^{-24} \text{ g.}$$

Radiation density constant :

$$a = 8\pi^5 k^4/15c^3h^3 = \left(\frac{V_0 A_0}{T_0}\right)^4 \frac{4\pi N_0 R_\infty (e/m)}{15 c^6 F^6}$$

$$= (7{\cdot}569_{42} \pm 0{\cdot}004_9) \times 10^{-15} \text{ erg. cm.}^{-3} \text{ deg.}^{-4}$$

Ratio mass ^{1}H atom to mass electron :

$$M_{^1\text{H}}/m = (e/m)({}^1\text{H}/F) = 1837{\cdot}5_{61} \pm 0{\cdot}5_6$$

* See long footnote in Section D. In my Aug. 1939 mimeographed list, the probable error of $R_0 T_0$ should have been 0·00008, in place of the stated 0·0008.

Ratio mass proton to mass electron

$$M_P/m=(e/m)\left(\frac{^1\mathrm{H}-\mathrm{E}}{F}\right)=1836{\cdot}5_{81}\pm0{\cdot}5_6$$

Second radiation constant:

$$c_2=hc/k=\frac{T_0c^2}{V_0q_0}\left\{\frac{2\pi^2F^2}{R_\infty N_0^2(e/m)}\right\}^{1/2}=1{\cdot}4384_9\pm0{\cdot}0003_4 \text{ cm. deg.}$$

Specific charge of α-particle:

$$2e/M_\alpha=\frac{2F}{\mathrm{He}-2\mathrm{E}}=4822{\cdot}3_3\pm0{\cdot}5_1 \text{ abs.-e.m.u. g.}^{-1}$$

Specific charge of proton:

$$e/M_P=\frac{F}{^1\mathrm{H}-\mathrm{E}}=9578{\cdot}7_7\pm1{\cdot}0 \text{ abs.-e.m.u. g.}^{-1}$$

Stefan-Boltzmann constant:

$$\sigma=ac/4=2\pi^5k^4/15c^3h^3=\left(\frac{V_0A_0}{T_0}\right)^4\frac{\pi^2N_0R_\infty(e/m)}{15(Fc)^3}$$

$$=(5{\cdot}672_{00}\pm0{\cdot}003_1)\times10^{-5} \text{ erg. cm.}^{-2} \text{ deg.}^{-4} \text{ sec.}^{-1}$$

Wave-length associated with one abs. volt *:

$$\lambda_0=10^{-8}c^2(h/e')=\frac{c^2}{10^8}\left\{\frac{2\pi^2F^2}{R_\infty N_0^2(e/m)}\right\}^{1/2}=(12395{\cdot}_4\pm2{\cdot}_1)\times10^{-8} \text{ cm. abs.-volt}$$

Wave number associated with one abs. volt:

$$s_0=1/\lambda_0=\frac{10^8}{c^2}\left\{\frac{R_\infty N_0^2(e/m)}{2\pi^2F^2}\right\}^{1/2}=8067{\cdot}4_9\pm1{\cdot}4 \text{ cm.}^{-1} \text{ abs.-volt}^{-1}$$

Wien's displacement-law constant†:

$$A=c_2/4{\cdot}965114=0{\cdot}28971_8\pm0{\cdot}00007 \text{ cm.deg.}$$

Zeeman displacement per gauss:

$$(e/m)/4\pi c=(4{\cdot}6699_1\pm0{\cdot}0013)\times10^{-5} \text{ cm.}^{-1} \text{ gauss.}^{-1}$$

REFERENCES

ANDERSON, W. C., 1937. *Rev. Sci. Instrum.* **8**, 239.

ANDERSON, W. C., 1941. *J. Opt. Soc. Amer.* **31**, 187.

BAXTER, G. P. and STARKWEATHER, H. W., 1926. *Proc. Nat. Acad. Sci., Wash.*, **12**, 699.

BAXTER, G. P. and STARKWEATHER, H. W., 1928. *Proc. Nat. Acad. Sci., Wash.*, **14**, 57.

BAXTER, G. P., GUICHARD, M., HÖNIGSCHMID, O. and WHYTLAW-GRAY, R., 1941. *J. Amer. Chem. Soc.* **63**, 845. Eleventh Report of the International Committee on Atomic Weights.

BATUECAS, T., 1935. *Boletin de la Universidad de Santiago*, Oct.-Dec. 1935.

BATUECAS, T. and CASEDO, F. L., 1936. *J. chim. Phys., Paris*, **33**, 41.

BEARDEN, J. A., 1931. *Phys. Rev.* **38**, 2089.

BEARDEN, J. A., 1935. *Phys. Rev.* **48**, 385.

BEARDEN, J. A., 1938. *Phys. Rev.* **54**, 698.

BEARDEN, J. A., 1939 a. *Phys. Rev.* **55**, 584.

BEARDEN, J. A., 1939 b. *Phys. Rev.* **56**, 1023.

BEARDEN, J. A., 1941. *J. Appl. Phys.* **12**, 395.

BEARDEN, J. A. and SCHWARZ, G., 1941. *Phys. Rev.* **59**, 934.

* The factor 10^{-8} was accidentally omitted, in G.C. 1929, in the equation for λ_0.

† The factor 4·965114 is the root of $e^{-p}+(p/5)-1=0$.

BEATTIE, J. A., 1941. *Temperature (Its Measurement and Control in Science and Industry)* (Reinhold Pub. Co.), New York, pp. 74–89.
BIRGE, R. T., 1929. *Rev. Mod. Phys.* **1**, 1.
BIRGE, R. T., 1932 a. *Phys. Rev.* **40**, 207.
BIRGE, R. T., 1932 b. *Phys. Rev.* **40**, 228.
BIRGE, R. T., 1932 c. *Phys. Rev.* **42**, 736.
BIRGE, R. T., 1934 a. *Nature, Lond.*, **133**, 648.
BIRGE, R. T., 1934 b. *Nature, Lond.*, **134**, 771.
BIRGE, R. T., 1935. *Phys. Rev.* **48**, 918.
BIRGE, R. T., 1936. *Nature, Lond.*, **137**, 187.
BIRGE, R. T., 1938. *Phys. Rev.* **54**, 972.
BIRGE, R. T., 1939 a. *Amer. Phys. Teacher*, **7**, 351.
BIRGE, R. T., 1939 b. *Phys. Rev.* **55**, 1119.
BIRGE, R. T., 1940 a. *Phys. Rev.* **57**, 250.
BIRGE, R. T., 1940 b. *Phys. Rev.* **58**, 658.
BIRGE, R. T., 1941. *Phys. Rev.* **60**, 766.
BIRGE, R. T. and JENKINS, F. A., 1934. *J. Chem. Phys.* **2**, 167.
BOLLMANN, V. L. and DUMOND, J. W. M., 1938. *Phys. Rev.* **54**, 1005.
BOND, W. N., 1930. *Phil. Mag.* **10**, 994.
BOND, W. N., 1931. *Phil. Mag.* **12**, 632.
DE BRAY, M. E. J. GHEURY, 1927 a. *Astr. Nachr.* **230**, 449.
DE BRAY, M. E. J. GHEURY, 1927 b. *Nature, Lond.*, **120**, 404 and 602.
DE BRAY, M. E. J. GHEURY, 1931. *Nature, Lond.*, **127**, 522.
DE BRAY, M. E. J. GHEURY, 1934. *Nature, Lond.*, **133**, 464 and 948.
CASIMIR, H. B. G., 1939. *Nature, Lond.*, **143**, 465.
CHRISTY, R. F. and KELLER, J. M., 1940. *Phys. Rev.* **58**, 658.
CHU, D. Y., 1939. *Phys. Rev.* **55**, 175.
CLEWS, C. J. B. and ROBINSON, H. R., 1940. *Proc. Roy. Soc.* A, **176**, 28.
CRAGOE, C. S., 1941. *Bur. Stand. J. Res., Wash.*, **26**, 495.
CURTIS, H. L., CURTIS, R. W. and CRITCHFIELD, C. L., 1939. *Bur. Stand. J. Res., Wash.*, **22**, 485.
CURTIS, H. L., MOON, C. and SPARKS, C. M., 1936. *Bur. Stand. J. Res., Wash.*, **16**, 1.
CURTIS, H. L., MOON, C. and SPARKS, C. M., 1938. *Bur. Stand. J. Res., Wash.*, **21**, 375.
DOLE, M. and SLOBOD, R. L., 1940. *J. Amer. Chem. Soc.* **62**, 471.
DRINKWATER, J. W., RICHARDSON, Sir OWEN and WILLIAMS, W. E., 1940. *Proc. Roy. Soc.* A, **174**, 164.
DUMOND, J. W. M., 1939. *Phys. Rev.* **56**, 153–64.
DUMOND, J. W. M., 1940. *Phys. Rev.* **58**, 457–66.
DUMOND, J. W. M. and BOLLMAN, V. L., 1936. *Phys. Rev.* **50**, 524.
DUMOND, J. W. M. and BOLLMAN, V. L., 1937. *Phys. Rev.* **51**, 400.
DUMOND, J. W. M. and BOLLMAN, V. L., 1938. *See* BOLLMAN and DUMOND, 1938.
DUMOND, J. W. M., BAILEY, H. H., PANOFSKY, W. K. H. and GREEN, A. E. S., 1941. *Phys. Rev.* **59**, 219.
DUMOND, J. W. M., PANOFSKY, W. K. H. and GREEN, A. E. S., 1941. *Phys. Rev.* **60**, 163.
DUNNINGTON, F. G., 1933. *Phys. Rev.* **43**, 404.
DUNNINGTON, F. G., 1937. *Phys. Rev.* **52**, 475.
DUNNINGTON, F. G., 1939. *Rev. Mod. Phys.* **11**, 65–83.
EDMONDSON, F. K., 1934. *Nature, Lond.*, **133**, 759.
VON FRIESEN, S., 1935. *Dissertation* (Uppsala).
GERLACH, W., 1926. *Geiger-Scheel Handbuch der Physik*, **22**, 48–82.
GOEDICKE, E., 1939. *Ann. Phys., Lpz.*, **36**, 47.
GREENE, C. H. and VOSKUYL, R. J., 1939. *J. Amer. Chem. Soc.* **61**, 1342.
GUILLAUME, CH.-ED., 1927. *La Création du Bureau internationale des Poidres et Mesures et son Œuvre* (Paris). See p. 258.

GUTTON, C., 1911. *C.R. Acad. Sci., Paris*, **152**, 1089.
HAHN, O., FLÜGGE, S. and MATTAUCH, J., 1940. *Phys. Z.* **41**, 1.
HARTSHORN, L. and ASTBURY, N. F., 1937. *Philos. Trans.* A, **236**, 423.
HENNING, F. and JAEGER, W., 1926 a. *Geiger-Scheel Handbuch der Physik*, **2**, 494.
HENNING, F. and JAEGER, W., 1926 b. *Geiger-Scheel Handbuch der Physik*. **2**, 491.
HEYL, P. R., 1927. *Proc. Nat. Acad. Sci., Wash.*, **13**, 601.
HEYL, P. R., 1930. *Bur. Stand. J. Res., Wash.*, **5**, 1243.
HOPPER, V. D. and LABY, T. H., 1941. *Proc. Roy. Soc.* A, **178**, 243.
HOUSTON, W. V., 1927. *Phys. Rev.* **30**, 608.
HÜTTEL, A., 1940. *Ann. Phys., Lpz.*, **37**, 365.
JAEGER, W. and STEINWEHR, H. v., 1921. *Ann. Phys., Lpz.*, **64**, 305.
JAEGER, W. and STEINWEHR, H. v., 1926. *Z. InstrumKde.* **46**, 105.
KAROLUS, A. and MITTELSTAEDT, O., 1928. *Phys. Z.* **29**, 698.
KINSLER, L. E. and HOUSTON, W. V., 1934. *Phys. Rev.* **46**, 533.
KIRCHNER, F., 1931. *Ann. Phys., Lpz.*, **8**, 975.
KIRCHNER, F., 1932. *Ann. Phys., Lpz.*, **12**, 503.
KIRCHNER, F., 1939. *Ergebn. exakt. Naturw.* **18**, 26–77.
LABY, T. H. and HERCUS, E. O., 1927. *Philos. Trans.* A, **227**, 63.
LABY, T. H. and HERCUS, E. O., 1935. *Proc. Phys. Soc.* **47**, 1003.
MANIAN, S. H., UREY, H. C. and BLEAKNEY, W., 1935. *J. Amer. Chem. Soc.* **56**, 2601.
MAYO, R. L. and ROBINSON, H. R., 1939. *Proc. Roy. Soc.* A, **173**, 192.
MERCIER, J., 1923 a. *Ann. Phys., Paris*, **19**, 248.
MERCIER, J., 1923 b. *Ann. Phys., Paris*, **20**, 5.
MERCIER, J., 1924. *J. Phys. Radium*, **5**, 168.
MICHELSON, A. A., 1927. *Astrophys. J.* **65**, 1.
MICHELSON, A. A., PEASE, F. G. and PEARSON, F., 1935. *Astrophys. J.* **82**, 26. Professor Michelson died in 1931, shortly after the measurements were started.
MILLER, P. H., Jr. and DUMOND, J. W. M., 1940. *Phys. Rev.* **57**, 198.
MITTELSTAEDT, O., 1929 a. *Ann. Phys., Lpz.*, **2**, 285.
MITTELSTAEDT, O., 1929 b. *Phys. Z.* **30**, 165.
MOLES, E., TORAL, T. and ESCRIBANO, A., 1939. *Trans. Faraday Soc.* **35**, 1439.
MURPHEY, B. F., 1941. *Phys. Rev.* **59**, 320.
OHLIN, P., 1941. *Dissertation* (Uppsala).
OSBORNE, N. S., STIMSON, H. F. and GINNINGS, D. C., 1939. *Bur. Stand. J. Res., Wash.*, **23**, 197.
PERLMAN, M. L. and ROLLEFSON, G. K., 1941. *J. Chem. Phys.* **9**, 362.
PERRY, C. T. and CHAFFEE, E. L., 1930. *Phys. Rev.* **36**, 904.
POINCARE, H., 1913. *The Foundations of Science* (New York : Science Press).
RAMSEY, A. S., 1940. *Nature, Lond.*, **146**, 180.
RAYLEIGH, Lord, 1881 a. *Nature, Lond.*, **24**, 382.
RAYLEIGH, Lord, 1881 b. *Nature, Lond.*, **25**, 52.
RIESENFELD, E. H. and CHANG, T. L., 1936. *Phys. Z.* **37**, 690.
ROBINSON, C. F., 1939. *Phys. Rev.* **55**, 423.
ROBINSON, H. R., 1938. *Nature, Lond.*, **142**, 159.
ROEBUCK, J. R. and MURRILL, T. A., 1941. *Temperature* (*Its Measurement and Control in Science and Industry*) (New York, Reinhold Pub. Co.), pp. 60–74.
ROSA, E. B. and DORSEY, N. E., 1907. *Bur. Stand. Bull., Wash.*, **3**, 433–604. See also pp. 605–22.
ROSS, P. A. and KIRKPATRICK, P., 1934. *Phys. Rev.* **46**, 668.
SCHEEL, K. and BLANKENSTEIN, F., 1925. *Z. Phys.* **31**, 202.
SHANE, C. D. and SPEDDING, F. H., 1935. *Phys. Rev.* **47**, 33.
SHAW, A. E., 1938. *Phys. Rev.* **54**, 193.
SMYTHE, W. R., 1934. *Phys. Rev.* **45**, 299.

STOTT, V., 1929. *Nature, Lond.*, **124**, 622.
SUYDAM, V. A., 1940. *Fundamentals of Electricity and Magnetism* (New York: Van Nostrand), pp. 616–627.
SWARTOUT, J. A. and DOLE, M., 1939. *J. Amer. Chem. Soc.* **61**, 2025.
VIGOUREUX, P., 1938 a. *N.P.L. Collected Res.* **24**, 173.
VIGOUREUX, P., 1938 b. *N.P.L. Collected Res.* **24**, 277.
VINAL, G. W. and BATES, S. J., 1914. *Bur. Stand. Bull., Wash.*, **10**, 425.
VINAL, G. W. and BOUVARD, W. M., 1916. *Bur. Stand. Bull., Wash.*, **13**, 147.
WENSEL, H. T., 1939. *Bur. Stand. J. Rt ., Wash.*, **22**, 375–96.
WILLIAMS, R. C., 1938. *Phys. Rev.* **54**, 568.

RECENT TELEVISION DEVELOPMENTS

By V. K. ZWORYKIN and R. E. SHELBY,
R.C.A. Manufacturing Corporation, Camden, N.J., and
National Broadcasting Company, New York, N.Y.

§ 1. HISTORICAL INTRODUCTION

ALTHOUGH this *Report* covers principally the most modern developments in the field of television, a brief historical outline of the inventions, discoveries and improvements which contributed most to its progress still serves a useful purpose. During the past several decades the technical advances in all branches of electrical communication—particularly radio—have been nothing short of phenomenal. It is therefore not altogether surprising that the development of electronic television is likely to be taken as a matter of course, and the fact that it had its humble beginnings often completely forgotten.

Probably the first practical discovery leading to the development of modern television was that which occurred in Valentia, Ireland, in 1873 (Larner, 1929), when a telegraph operator named May noticed that his instruments behaved erratically when the sun shone through the window on his selenium resistors. The photo-conductive properties of selenium, based on the observations of May, were reported to the Society of Telegraph Engineers by Willoughby-Smith in the same year. This discovery aroused considerable interest and stimulated inventors to propose, within the next few years, methods of picture transmission involving this principle. The next milestone in the development of modern television, although it was not recognized as such at the time, was the invention in 1878 of the Crookes tube by Sir William Crookes and the demonstration of the properties of cathode rays. However, it remained for Paul Nipkow to introduce the concept of scanning with his invention, in 1884, of a system involving a rotating disc, having apertures arranged in spiral fashion, a photo-sensitive element, and synchronized reproducing means. The sluggishness of selenium cells, although inventors apparently did not realize it at the time, rendered them impractical for any high definition television system, and the discovery of the photo-electric effect by Hertz in 1888, eventually leading to the development of vacuum photo-cells having negligible inertia, therefore ranks as the next important milestone.

In 1907, Boris Rosing in Russia and A. A. Campbell-Swinton in England separately and simultaneously published methods of electrical image reproduction using electromagnetic means for scanning. The first all-electronic television system, utilizing cathode-ray scanning at both transmitter and receiver, was

later proposed by Campbell-Swinton in 1911 in an address before the Roentgen Society. In many respects this proposal visualized the present-day systems, but it lacked one important factor. That factor was storage of the electric charge between successive scannings—the principle that would give the all-electronic pickup device the sensitivity necessary to make it practical with high-definition scanning. Unfortunately, a practical demonstration of these early proposals was impossible at the time, chiefly because the epochal invention of the thermionic amplifier tube by De Forest in 1907 still needed further development to make practical video amplification available.

In the next 25 years or so, various improvements were suggested, but it seems that none of these proposals reached the demonstration stage until the early 1920's. In 1923, Mr. John Logie Baird in England and Mr. C. Francis Jenkins in the United States gave successful demonstrations of moving silhouetted figures. Both these inventors used electro-mechanical scanning methods. It was also in 1923 that V. K. Zworykin (1938) applied for patents on his iconoscope, a cathode-ray television camera tube having inherently greater sensitivity than any previously proposed, by virtue of its property of storage over relatively long periods of time of the electric charges resulting from photo-electric emission.

During the next 10 or 15 years, extending into the middle 1930's, electro-mechanical television methods reached the highest peak of their development, culminating in the use of the electron multiplier type high-vacuum photo-cells for maximum signal-to-noise ratio at the sending end.

In the early 1930's, P. T. Farnsworth (1934) with his dissector tube and V. K. Zworykin (1929, 38) with the iconoscope demonstrated television systems of fairly good definition, based exclusively on electronic scanning. By 1936, the all-electronic method, corresponding to the demand for television images of higher and higher definition, had so far outdistanced mechanical scanning methods that the prominent experimenters in the field discarded mechanical scanning entirely, except in a few instances for the televising of film. From 1936 to date, the demand has been for further refinements and higher definition, culminating in the 525 line 30 frame interlaced images being broadcast commercially in the United States to-day.*

Storage type cathode-ray camera tubes, because of their higher sensitivity, are now universally used for direct pickup purposes. Although the Farnsworth dissector tube, being of relatively low sensitivity, is no longer used for direct pickup, it has been found to have useful applications in the transmission of motion-picture film (Jensen, 1941). With one or two notable exceptions (Robinson, 1939; D. G. F., 1940), television equipment to-day is entirely electronic, involving no mechanical moving parts.

* In August 1936, high definition (405 line, 25 frame interlaced) television transmission was also initiated by the British Broadcasting Corporation (see Macnamara and Birkinshaw, 1938)—[EDITOR].

§ 2. PROBLEMS OF MODERN TELEVISION

(a) *Pickup*

Television programs for home entertainment are derived from three main sources—direct pickup from a studio, outside pickup from the scene of an event, and pickup from film. In order better to understand the development of the instrumentalities used in television broadcasting, it will be helpful to review briefly the requirements of each of these three types of pickup (Macnamara and Birkinshaw, 1938; Morris and Shelby, 1937; Protzman, 1940; Hanson, 1937; Lohr, 1940; Fink, 1940; Zworykin and Morton, 1940).

Studio. The requirements to be met in designing a direct-pickup television studio centre around the fundamental fact that a television program is continuous, as distinguished from motion-picture production. Once a television program has begun, the action must be continued without interruption to its end, except for brief intermissions in long productions. This is in contrast to the technique that may be employed in motion-picture production, where a single scene or a single sequence may be photographed after elaborate preparation, and any desired amount of time taken before filming the following sequence. This leads to the requirement that the television studio must be large enough to accommodate a minimum of two scenes and preferably several scenes. It would, of course, be possible to have successive scenes televised from different studios, but experience has shown that in most cases this is less desirable than the use of a single studio. The studio must be large enough to accommodate not only the sets of the scenes to be employed, but also the television cameras, microphone stands and booms, lighting equipment, and other auxiliary apparatus associated with the television operation. Usual practice calls for a control room immediately adjacent to the studio, where the programme producer and control operators can watch the studio action through a window and at the same time monitor the televised scenes on television screens (Hanson, 1937). In order to clear the floor of the studio of apparatus not absolutely necessary, and also to provide better lighting, most of the lighting fixtures are generally suspended from the ceiling. This requires a studio with a fairly high ceiling in order to get suitable spacing of the lighting fixtures from the scene to be illuminated and also to prevent the fixtures from showing in the televised scene. The high ceiling is also an aid in keeping the temperature of the air in the studio down to a satisfactory level.

The sound pickup which accompanies all television programs must be provided for, and certain problems arise in television operation which are not normally encountered in sound broadcasting. The sound-pickup technique in television operation is rather more like the sound-pickup technique in motion-picture production than in sound broadcasting, but again there is the important difference that action must be continuous. Because of the fact that it is usually desired that the television audience shall not see the microphone, it is often located farther from the performers than is the case in sound broadcasting.

This means that the acoustic properties of the studio take on added importance and must be somewhat different in television from what they are in sound broadcasting. Experience has shown that television studios should be designed to have a slightly shorter reverberation time than would be considered optimum for the same studio if used for sound broadcasting. This is due in part to the previously mentioned fact that the pickup microphone is, in the average case, farther from the performers than in sound broadcasting, and in part to the fact that the scenery, furniture and equipment in the television studio modify the acoustic properties in a way that tends to liven the pickup (Hanson, 1937).

In order to provide the best possible sound pickup and still keep the microphone out of the scene, television has adopted the same type of boom for the microphone as is used in sound-motion-picture operation. The boom consists of a multiple section telescoping arm, mounted on a movable stand with controls which permit an operator to follow the action in the television scene with the microphone on the end of the telescoping arm, always just above the scene but near enough to the performers to obtain satisfactory sound pickup.

Lighting presents one of the most difficult problems in television operation. The amount of light required for satisfactory results with the usual types of scene is dependent upon the sensitivity of the television pickup device and upon the light-gathering powers of the lens associated with it. As in photography, if the aperture of the lens is made too large, then the depth of focus is very limited and results are unsatisfactory. With television-pickup tubes of the iconoscope type, which is at present the most widely used type, satisfactory results are obtained if the lenses are operated with apertures which give speeds in the range from f/3·5 to f/5·6 and incident light on the scene ranges from 300 to 1500 foot-candles, depending upon the nature of the scene being televised. With recently developed pickup tubes of higher sensitivity, such as the orthicon, the above light requirements may be substantially lowered.

In lighting a scene, either for photography or television, the amount of light provided is only one of a number of requirements if pleasing results are to be obtained. The direction, diffusion and spectral quality of the light all have an important bearing upon the results obtained. At the present time, incandescent lamps are used more extensively for television lighting than other light sources, but the carbon arc provides very satisfactory lighting and many other types of light sources are being tested experimentally. All the knowledge and techniques developed in lighting for still photography and motion pictures find important application in television lighting, except that the television system is not yet able to reproduce the wide range of light shades possible in photography, and for this reason it is not possible to employ successfully certain of the dramatic lighting techniques which call for extremely harsh contrasts between highlights and shadows.

Television-lighting technique differs from that employed in stage lighting or in motion-picture production, and again the differences are due mainly to the

requirement of continuous action. It is true that continuous action is also required in the theatre, but here the audience's viewpoint is fixed—that is, the physical relation of the audience to the scene remains the same throughout the play. In a motion picture, this relation may not remain fixed, but in this case the requirement of continuous action in photographing the show does not obtain. In television production, on the other hand, if more than one camera is used, or if a single camera is moved about, then the viewpoint of the audience is shifted and the lighting must be planned accordingly. It must not favour the pickup of one camera in one position to the detriment of the scene as viewed from other angles. As it is not feasible to provide instantaneous shifts in lighting when switching from one camera to another, the lighting of a given scene must represent a compromise between the requirements for all cameras, although some manipulation of the lighting is possible during the course of the program, if suitable flexibility is provided in the design of the lighting fixtures. A type of fixture for overhead use in television studios which allows adjustment of the direction of the light output from each fixture by means of remote control from a central point has been developed and described in the literature (Eddy, 1940).

The scenery used in television-studio productions must also meet certain special requirements, and here again the requirement for continuous action has its effect. The design of stage scenery is simplified by the fact that, roughly speaking, the scene is viewed from a single direction. This means that in many cases the scenery may be arranged in parallel rows facing the audience. Motion-picture technique, on the other hand, must provide a third dimension in most cases, since the viewpoint may be changed by moving the camera. This is also true in television, with the added requirement that, generally, " time out " may not be taken to rearrange the scenery for successive shots, as is the case in motion-picture production. These special requirements may be summed up by stating that the scenery must give the proper illusion when viewed from all camera positions. Obviously, compromises are necessary, since the perspective will be correct in a painted scene only when viewed from the proper angle. It might be thought that television would suffer a serious handicap because of this, but experience has shown that a clever scenic artist can produce completely satisfactory results for any reasonable scene with the co-operation of the producer in the matter of carefully planning the angles of the various camera shots.

It has been found helpful in the painting of television scenery to avoid the use of colour and to paint all scenery in various shades of grey. This minimizes the difficulties which might be encountered due to slight differences in spectral response of the pickup tubes in different cameras. Limited depth of focus, due to the relatively fast lens employed in the television camera, results in a slight defocusing of most of the scenery in the television set, if the performers are in sharp focus. Because of this it has been found helpful in the painting of scenery to employ somewhat sharper, harsher lines than would be used if the scenery were intended for direct viewing by eye. The softening effect of the slight

defocusing then results in a scene which is not quite as blurred as would otherwise be the case.

An important requirement for television scenery is that it must be demountable quickly and silently, so that one scene in the studio may be rearranged while others are being televised, without objectionable noise being heard by the audience. The type of scenery most often used consists of cloth mounted on wooden frames and supported by fixtures of the type widely employed for theatrical scenery.

In addition to the ordinary painted scenery previously referred to, some use has been made in television of scenery projected upon a translucent screen behind the performers. This is a process which has been used widely both in still photography and motion-picture production. The somewhat higher light levels used in television call for a very bright projected picture, and this in turn limits the process at the present time to fairly small projected areas, unless extremely powerful light sources are employed. Another limitation is that the projected scene is of satisfactory brilliance only when viewed within a relatively narrow angle in front of the screen in the case of screens that are practical for this use (Protzman, 1940).

It was mentioned previously that the programme producer and technical control staff are generally located in a room separated from, but immediately adjacent to, the studio proper. This facilitates their work by allowing a view of the studio as well as of the television-monitoring screens, and at the same time permits them to talk freely with one another and by telephone to other points concerned with the television broadcast without interfering with the sound pickup in the studio. In order to afford an unobstructed view of the action, the control room is generally located somewhat above the studio floor level (Morris and Shelby, 1937).

In addition to the picture monitors in the control booth, the television amplifiers, scanning amplifiers, power supplies and control equipment associated with the studio cameras are usually located in the control booth (Morris and Shelby, 1937; Protzman, 1940). A monitoring loud-speaker and controls for regulating sound level and for fading between microphones are also located in the control room, and a sound-control engineer is provided in the booth to operate them. Controls for regulating picture brightness and contrast and for switching between cameras in the studio, as well as various other controls associated with the television equipment, are also located in the booth and are operated by control enginers. The camera operators and the operators of the microphone booms, as well as certain other operating personnel in the studio, all wear headphones, by means of which instructions may be received from the programme producer and engineers in the control booth (Morris and Shelby, 1937).

Outside pickup. Some of the most interesting television programmes consist of the televising of events as they occur outside the studio. This corresponds roughly to news-reel coverage in motion pictures, but it has the added appeal of immediacy—that is, the audience does not know the outcome of the event as it

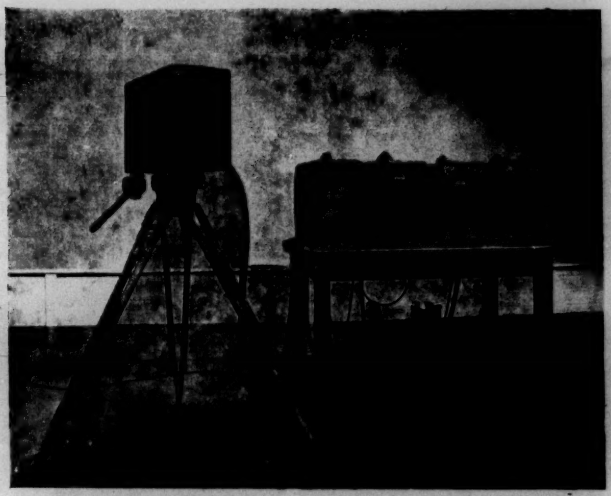


Figure 1.

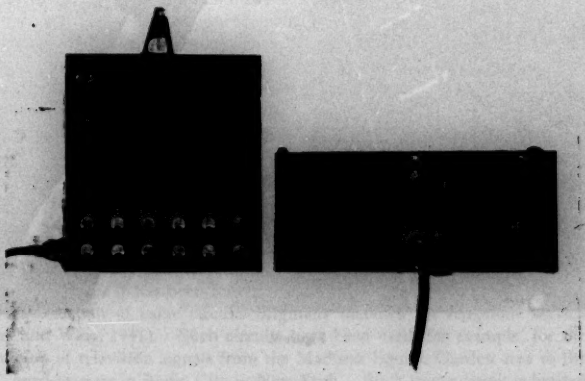


Figure 2.

To face page 140.

Figure 9.

watches it. Many people believe that television will find its most important application in this field. At any rate, it seems fairly certain that the extension of the television "eye" to points outside the studio is even more important to television than was the extension of the microphone to such outside pickup points in sound broadcasting (Hanson, 1941).

In order to provide on-the-scene remote television pickup, equipment must be provided which is the equivalent of a complete studio plant, plus some facility for relaying the signal back to the main plant or the broadcast transmitter. The problem is analogous to that of remote pickup work in sound broadcasting, but the equipment required is many times more complicated, and, in fact, must include complete facilities for sound pickup in addition to those for picture pickup. The most direct and complete solution to this problem has been the utilization of equipment similar to that employed in studio installations, mounted in trucks large enough to accommodate this relatively bulky and heavy equipment. Two such trucks have been employed by the N.B.C. for this purpose for several years past in the New York City area. One of these trucks houses the equipment associated with the cameras, the other a radio transmitter for relaying the signal back to the main point.

Another solution to the outside pickup problem has been provided in the form of equipment built in units small enough to be carried, which may be taken to the scene of the pickup by any suitable means of transportation and set up in the most advantageous position. In this case there has been some sacrifice in performance and operating flexibility for the sake of portability, but, nevertheless, such equipment is capable of providing high-quality television pictures and has the advantage of portability, which makes it possible to televise many scenes that might not be accessible to the equipment mounted permanently in large trucks (Beers, Schade and Shelby, 1940; Smith, 1940; Hanson, 1941).

The photograph of figure 1 illustrates typical portable television pickup equipment, and that of figure 2 shows a portable radio relay transmitter of complementary design.

The problem of relaying the television signal from the point of pickup to the main plant or broadcast station is an important one in remote television pickup work. In the case of sound broadcasting, it is possible to utilize wire lines which have been installed for telephone service and which are widely available. Circuits designed for the transmission of sound signals are, however, useless for the relaying of television signals on account of the wide band of frequencies associated with the latter. For transmission of television signals over short distances it has been found possible to obtain satisfactory results by special equalization of cable circuits originally installed for telephone service (Strieby and Weis, 1941). Such circuits have been used, for example, for the transmission of television signals from the Madison Square Garden area to the N.B.C. headquarters in Radio City in New York. Such transmission is limited to circuit lengths of approximately one mile between repeaters, and if the distance

involved is more than a few miles, it would appear that other transmission means are more desirable. Since wide band circuits do not exist except in a very few special installations, most of the television-relay work has been done by means of radio circuits. In its television service in the New York area the N.B.C. has utilized mainly two radio transmitters for this purpose. One of these operates in the television channel 162 to 168 megacycles and is mounted in one of the trucks mentioned above. The distance over which it is possible to operate such a relay is dependent upon many variable factors, such as antenna height, interference at the receiving point, and terrain between the two points. Satisfactory programme transmissions have been achieved over a distance of 26 miles, utilizing the transmitter operating in the channel 162 to 168 megacycles. It is possible to extend this distance by multiple relays, and in this manner a television programme was successfully handled at a pickup point 68 miles from the main plant, utilizing three radio transmitters in a relay link service. As discussed elsewhere in this *Report*, an extension of this principle makes possible television-network operation.

An outstanding example of a wire line facility for long-distance television transmission is the coaxial cable installation of the American Telephone and Telegraph Company between New York City and Philadelphia (Espenschied and Strieby, 1934). This has recently been used for a number of television demonstrations and experiments, including the outstanding series of programmes from the scene of the Republican Party National Convention in Philadelphia during June 1940 (Hanson, 1941). Scenes from the floor of the convention hall were televised by the Mobile Units of N.B.C., transmitted to New York via the coaxial cable and connecting facilities of the A.T. and T., and broadcast in New York City by the N.B.C. television transmitter. As its name implies, the coaxial cable consists of a central conductor surrounded by an outer conductor having, in this case, a diameter of approximately 5/8 inch and a characteristic impedance of approximately 72 ohms (Wentz, 1940). This cable has an attenuation of approximately 6 db. per mile at a frequency of one megacycle, and the installation between New York and Philadelphia, as it was used for the programme mentioned above, had repeater amplifiers approximately every five miles throughout its entire length in order to keep the signal sufficiently above the noise level on the circuit.

Film pickup. Any programme or scene which may be recorded on motion-picture film may be televised subsequently from the film with modern television facilities. This, of course, opens up tremendous programme possibilities for a television service. It is true that the overall resolution capabilities of even the best television system is not good enough to reproduce the finest resolution which is possible on the best motion-picture film. Experience has shown, however, that it is good enough to be acceptable in any case, and in most cases the loss in definition is not noticed by the average viewer, unless a direct comparison is made.

In addition to its importance as an independent source of programme

material, television pickup from film is useful as an auxiliary service for direct pickup television. In many cases certain scenes in dramatic productions cannot be provided in a studio of reasonable size, or at the time the programme is televised. Such scenes may be photographed on motion-picture film and televised at the proper time in the programme to give the desired continuity. Another important application for film in television broadcasting is in the programming of important events which occur at a time when most of the potential audience is at work. Such programmes may be photographed on film in the afternoon, for example, the film processed and edited, and transmitted by television the same evening when the maximum audience is available.

Years of experience in the motion-picture industry have shown that the most satisfactory type of motion-picture projector for general use is the intermittent type. A special problem arises in the use of the intermittent projector in television because of the difference in standards between the motion-picture and television industries. In the case of motion pictures, the standard frame-repetition rate is 24 frames per second, while in television the frame-repetition rate is usually a sub-multiple of the power-supply frequency. Thus, for example, in the United States, the frame-repetition rate has been standardized at 30 frames per second, since the most widely used power-supply frequency is 60 cycles per second. In England, and on the Continent, the television standard is 25 frames per second, since the most widely used power-supply frequency is 50 cycles per second. In the latter case, if standard motion-picture films are projected at a speed of 25 frames per second, instead of 24 per second, the results are generally acceptable, although there is a slight distortion in sound reproduction which is noticeable on certain types of programmes. In the case of the U.S. standard, the operation of standard motion-picture film at 30 frames per second would produce intolerable distortion in the sound and serious distortion in the motion of objects in the reproduced scene. It would be technically possible to use only films which were taken at the 30 frame per second rate, but this would be a serious handicap, since the vast libraries of standard film could not be used. It is possible to utilize film in a continuous projector operating at the standard film rate in conjunction with a television system operating at 30 frames per second, provided the continuous projector gives sufficiently good performance, and this is done in some instances (Jensen, 1941).

A completely satisfactory solution to the problem created by the difference in standards between the motion-picture industry and the television industry has been found which permits the use of film projectors of the intermittent type. This method depends upon the storage properties of the iconoscope pickup tube. These storage properties enable the iconoscope to " remember " for an appreciable time a picture which is focused on its light-sensitive mosaic and then removed. When the light image is impressed upon the mosaic, electrical charges are built up on the mosaic corresponding to the light intensity at each point. This electrical charge will remain after the light image is removed until

it is equalized by the scanning beam, provided the time interval is not long enough to allow the charge to leak off through imperfect insulation. This property of the iconoscope enables the usual film-projection cycle to be reversed —that is, instead of the shutter on the projector being closed for as short an interval as possible, during which time the film is pulled down from one frame to the next, the shutter is open only during the short interval when the scanning beam in the iconoscope is returning from the bottom to the top of the picture and the film pull-down occurs during the long interval when the shutter is closed and the mosaic is being scanned by the electron beam. In other words, the television picture is scanned with the iconoscope mosaic in darkness, the iconoscope " remembering " the picture which was projected on it during the preceding frame-blanking period.

Since interlaced scanning is standard and since, therefore, the scanning beam must be returned from the bottom to the top of the picture 60 times per second, it follows that the shutter is opened 60 times per second. The intermittent action in the projector is arranged so that alternate frames of the film remain in position in front of the aperture 50 % longer than the intervening frames. More specifically, one frame will remain in the aperture for two successive openings of the shutter, the third frame for two successive openings of the shutter, and so on. Since the shutter opens 60 times per second, it therefore follows that a complete operating cycle of the intermittent mechanism in the projector represents 2/60 plus 3/60, or 5/60 of a second. Thus one cycle covers the projection of two frames of the film. Therefore it is apparent that the average rate of travel through the projector is 24 frames per second, which is standard. This method of translation from a 24-frame per second rate for the film to a 30-frame per second rate for the interlaced television scanning is possible by virtue of the numerical relation of the two rates and the storage properties of the iconoscope pickup tube.

In a practical layout for television-programme operation it has been found desirable to have the film projectors located in one room and the television cameras and associated control equipment in an adjacent room (Hanson, 1937). When 35-millimetre film projectors are employed, the projection-room construction is usually required to meet certain rather stringent fire rules with regard to safety shutters, wall construction, exhaust vents, film-storage vaults, and so forth. These requirements are more readily met if the projectors are in a separate room from the other equipment. Another advantage of this arrangement is that the noise of the projectors does not interfere with monitoring in the control room. Usually two or more projectors are provided, so that continuous film operation is possible. This is further facilitated by the use of at least two television cameras. If the two cameras are movable, they may be used in conjunction with a number of projectors. In the N.B.C. film studio in Radio City, the two television cameras are mounted on tracks on the wall so that they may be moved in front of any one of several projection ports.

Television cameras employed for direct pickup have a lens system on the front of the camera for focusing an image of the scene to be televised upon the light-sensitive surface of the pickup tube. This arrangement is quite similar to that employed in cameras used for motion-picture and still photography. However, in the case of television cameras used with film projectors, it is customary to project the image upon the light-sensitive surface of the pickup tube by means of the objective lens mounted on the film-projection machine.

(b) *Amplification and distribution*

The signal output from any of the known pickup devices, at the present time, represents only a minute amount of power—no more than a few thousandths of a microwatt—and a voltage level of the order of a few millivolts. Before this signal can be used to modulate a radio transmitter of even moderate power it must be amplified to a level of several hundred watts. Thus the power gain of the chain of amplifiers raising the signal level must be of the order of 10^{12}. In a simple laboratory set-up, employing standard tubes, this gain might be realized by approximately twenty stages. In elaborate television-broadcasting plant, two or three times this number may be used, because at many points in the system it is necessary to make up for attenuation losses in coaxial cables, losses incident in controlling and shaping the signal, and so forth (Zworykin and Morton, 1940).

Video amplifiers, like those employed for audio-amplification purposes, make use of thermionic vacuum tubes. However, because of the much more exacting requirements imposed by the amplification of the video signal, the means used to couple the tubes are somewhat different. Because the general amplifier problem has been so ably and completely discussed in numerous articles (Foster and Rankine, 1941; Wheeler, 1939 a, 39 b; Bedford and Fredenhall, 1939; Herold, 1938) and textbooks (Fink, 1940; Zworykin and Morton 1940), it is unnecessary to discuss the entire theory of the amplifier here. Instead, the differences in the requirements of audio and video amplifiers will be briefly outlined.

The principal differences which distinguish the video amplifiers from the audio amplifiers are:—

1. A constant frequency response over a band several megacycles wide (a minimum of approximately 30 cycles to 4·5 Mc. for present U.S. standards).
2. An essentially constant time delay over the usable frequency range.

The resistance-coupled amplifier most nearly satisfies the above requirements —although it is by no means ideal—and it is the type that is used almost universally for video amplification. With the tubes available at present, it is not feasible to design resistance-coupled amplifiers which meet the requirements specified without resorting to some circuit arrangement which equalizes the effect of

unavoidable capacity across the coupling resistance. Even with a circuit design which reduces stray circuit capacity to a negligible quantity, the unavoidable inter-electrode capacities of the vacuum tubes still shunt the coupling resistance, and if this resistance is of a high enough value to provide a practical amount of voltage amplification, this combination resistance-capacitance network will have a frequency characteristic which falls off at the higher frequencies in the video range. A simple compensation for this effect is obtained by the addition of an inductance in series with the coupling resistor (see figure 3), the inductance being of such value that it will resonate with the capacity at a frequency somewhat above the highest frequency to be transmitted. In such a circuit, with capacity fixed, it is possible to choose inductance and resistance values such that the amplifier response will be essentially uniform over the video-frequency band (Herold, 1938).

This problem of providing uniform response over the video-frequency band may be considered as a low-pass filter network problem, and a very considerable

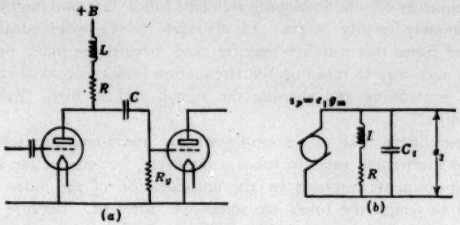


Figure 3.

amount of work has been done in applying classical filter theory to its solution (Wheeler, 1939 a, 39 b).

In addition to the problem at high frequencies in the video range, difficulty is encountered at the extremely low frequencies. Practical circuit considerations require the use of a blocking condenser between successive stages to isolate the different d.c. electrode potentials applied to the two tubes. At high frequencies this condenser will represent a negligible series impedance, even if it is of relatively low capacity, but at the extremely low frequencies its impedance becomes a problem. If it is made too large, its physical size may result in a capacitance to ground shunting the coupling impedance, which will affect the high-frequency response. Several practical expedients have been evolved to solve this problem. The diagram of figure 4 shows one such circuit arrangement. In this arrangement, an R.C. network is placed in series with the main coupling impedance in the plate circuit to give added response at low frequencies which will approximately compensate the loss at these frequencies caused by the coupling condenser and shunt grid resistor. Various other circuits have been proposed and are described in the literature.

The response of only a single stage of amplification has been considered so far. In practice, the complete amplifier consists of a great many cascaded stages, and the overall response is the important consideration (Bedford and Fredenhall, 1939). If the amplifier consists of identical stages and there is no incidental regeneration resulting from improper shielding, the overall gain will

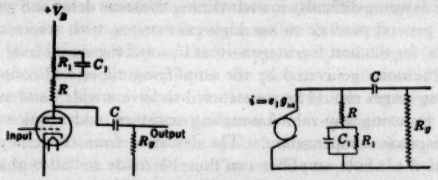


Figure 4.

be the gain of a single stage at each frequency raised to a power equal to the number of stages, while the time delay will be the product of the delay for each stage and the number of stages. The amplitude and delay curves for 16, 32 and 64 stages of amplification coupled as shown in figure 3, for two values of compensating inductance near the optimum value, are given in figure 5.

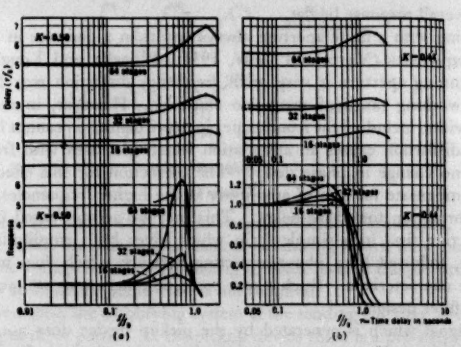


Figure 5.

It is often advantageous to employ amplifier stages which are not identical but which differ in such a way that they tend mutually to compensate their individual defects. Thus, a filter-coupled stage which has a flat response but a positive time delay error over part of the useful frequency band can be used with another constant-gain stage which has a negative delay error over the same portion of the band. If the characteristics of the individual stages differ, the

overall gain is the product of the gains at each frequency, and the total delay error will be the algebraic sum of the individual errors.

In amplifiers having a great many stages, as at the pickup, monitoring and modulating units of a transmitter, it is ordinarily not practical to make use of the complex coupling circuits required for high-gain stages throughout. The reason for this is the extreme difficulty in maintaining constant delay and gain characteristics. It is general practice to use high-gain stages, with response corrections in later stages, for the first few stages—that is, until the signal level is well above the level of the noise generated by the amplifying tubes and coupling circuits. The remaining stages should be constructed to have a wide band at the expense of efficiency by using low-valued coupling resistors, and require only a small amount of response compensation. The deviation from constancy of gain and time delay of the whole amplifier can thus be made to fall within acceptable limits.

There is an important consideration which modifies the previously given requirement of uniform amplitude response in video amplifiers over the entire band of signal frequencies. All pickup devices at present in use in practical television systems give a response which is more or less dependent upon frequency. This is due in part to the electrical characteristics of the devices and in part to aperture effects. In order to obtain undistorted image reproduction it is necessary that the overall response be flat.

Scanning with a finite aperture always results in a decrease in signal at the higher frequencies (Mertz and Gray, 1934; Wheeler and Loughren, 1938). If the scanning aperture is very small, frequency distortion may be negligible over the working range of the video amplifier. However, in most practical pickup devices, including the iconoscope, aperture distortion cannot be neglected. Aperture distortion causes an attenuation which increases with frequency but produces no change in phase delay. The correction for this effect, therefore, should compensate for loss of amplitude at the higher frequencies, but should give no correction for phase delay. This may be accomplished by providing the correction first in a simple form which gives both amplitude and phase correction, followed by a phase-correcting network which does not affect the amplitude characteristic. Such phase-correcting networks are well known in the art of filter design.

The signal which is generated by the pickup device does not contain the complete information necessary for the reconstruction of the picture. In addition to voltage variations representing the light and shade in the picture, it is necessary to furnish the information required to insure synchronism between the scanning pattern at the transmitter and the receiver. This is supplied in the form of properly shaped pulses, timed so that they occur after the end of each scanning line and before the beginning of the next. Similarly a pulse or series of pulses is added between field repetitions to insure vertical synchronization (Bedford and Smith, 1940).

The synchronizing impulses are added to the picture signal in the amplifier chain. This operation is performed by rendering the amplifier inactive ahead of the point of injection for a period equal to or slightly in excess of that occupied by the actual synchronizing signal, and injecting the impulse during this interval. The purpose of this " blanking " process is to prevent any possible signal from the pickup device from interfering with the synchronizing signal. During the blanking period the voltage level in the amplifier is brought to a value which corresponds to maximum black of the picture, thus providing a reference point from which the average illumination in the picture is indicated. If the pickup device is essentially an a.c. device, indicating only the variation in light and dark over the picture but not giving information as to the average illumination, as, for example, an iconoscope, then a special photo-tube and d.c. amplifier may be necessary to establish this value. The various steps in the formation of the complete signal are indicated in figure 6. In some pickup devices, such as the

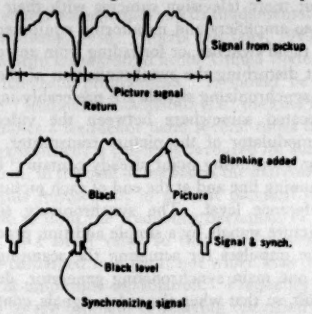


Figure 6.

orthicon (Rose and Iams, 1939 b) and image dissector (Farnsworth, 1934), the voltage level, when the scanning is blanked out during return time, bears a fixed relation to the voltage level corresponding to black level in the picture, and with such devices an auxiliary photo-tube is not required.

For convenience, the amplifying system at the sending end may be considered as consisting of three parts. At the pickup device there is a pre-amplifier which receives the very weak signal, usually of the order of a millivolt or at most a few millivolts, and raises it to a level of perhaps a tenth of a volt. This amplifier is made to have as low a noise level as possible, because the input at this point is not greatly in excess of the statistical fluctuations in the necessary coupling resistors and vacuum tubes. In order to obtain a maximum signal-to-noise ratio, it is necessary to make the gain of the first few stages as high as possible. This may be done even at the expense of loss of high frequencies in these early stages, which is made up by compensation in later stages.

The output of the pre-amplifier is arranged to drive a low-impedance cable (usually a low-loss coaxial) which connects it with a second amplifier chain, variously called the monitoring amplifier or the control amplifier. This unit raises the signal to a level of 1 to 10 volts. Also the gain, the d.c. level, etc., are controlled in this amplifier, and any special signals which may be necessary are added.

The output of this amplifier is fed to the line amplifier, which raises the signal level to that required by the modulator of the radio transmitter.

This subdivision of the amplifier chain into three, and only three, physical units is, of course, a simplification which does not exist in practice because the physical layout makes more than three units necessary. For example, in a typical television pickup and broadcast plant there are usually several studios, one of which is arranged to produce video signals from motion-picture film (Morris and Shelby, 1937; Hanson, 1937). Each studio will in general be equipped with two or more television cameras with their associated scanning wave generators, video amplifiers and monitoring equipment. Means must be provided for making rapid switches, or for fading from any one of these cameras to any other, without disturbing the synchronization of the television receiver. For this reason, the synchronizing signals are preferably inserted at a common monitoring point located somewhere between the video signal generating equipment and the modulator of the picture transmitter. The video signals reaching this common monitoring point already contain " blanking pedestals " at the end of each scanning line and at the end of each picture field, representing " picture black " reference level. The synchronizing signals can thus be combined with the picture signals by a simple addition process. The blanking signals, as well as the impulses for actuating the scanning amplifiers in each studio, are fed from one main synchronizing generator, delay networks being inserted in each circuit, so that when the video signals, containing the blanking pedestals, arrive at the common monitoring point from any camera or studio, they are delayed by the same amount and synchronizing signals can be added properly. Thus, any switching or fading between cameras or studios cannot upset the synchronizing signals or cause a loss of synchronism at the receiver. Therefore, the actual physical layout of the television plant is far more complicated than the idealized case of the three units already referred to, especially when provision is made to switch between studio subjects and remote pickups which are brought in to the common monitoring point via wire or radio links. Functionally, however, the three units can usually be easily identified.

(c) *Transmission and relay*

The design problem presented by the radio transmitter to be used for a television system is a formidable one (Blumlein, Brown, Davis, Green, 1938; Macnamara and Birkinshaw, 1938; Engstrom, 1939; D. G. F., 1939; Goldmark, 1939; McIlwain, 1939; Mertz, 1939; Sarnoff, 1939; Zworykin and Morton,

1940). This is because of the tremendous width of the band of frequencies which must be radiated if high picture definition is to be attained.

Like a sound transmitter, the transmitter in question may be made up of a carrier generator which produces a high-frequency oscillation of constant amplitude, and a modulator which varies the amplitude of this radio-frequency carrier in such a way that it is at every instant proportional to the signal applied to the modulator—in this case the video signal. The modulated radio frequency is then fed to the transmitting antenna, from which it is radiated. This is the general type of transmitter which has been most widely used in television work.

Recently proposals have been made for, and experiments conducted with, two other types of television transmitters. In one of these, the radio-carrier frequency is modulated instead of its amplitude, the frequency at any instant being governed by the voltage of the composite picture and synchronizing signal to be transmitted. This is a form of modulation which has received considerable attention recently in sound transmission and its fundamentals have been presented in the literature (Armstrong, 1936; Morris and Guy, 1940; Seeley, 1941). It is the type which has been standardized in the United States for the sound signal associated with the television programmes. It should be noted, however, that whereas " wide band " frequency modulation (i.e., the kind in which the complete signal spectrum occupies a frequency band several times the band width of the audio signal being transmitted) is generally favoured for such frequency modulation in sound operation, in the case of television the extremely wide band of video frequencies makes this inexpedient in present practical cases. Instead it has been found advisable to limit the frequency deviation, in frequency-modulated video transmission, to a small value so that the essential components of the signal spectrum are contained in a band not wider than would be required for the equivalent case of amplitude modulation. Frequency modulation has been applied successfully to the problem of relaying television signals from point to point where two or more stations may be necessary to cover the distances involved. In this application, the carrier frequencies employed have been somewhat higher than heretofore used, being in the vicinity of 500 Mc. (Kroger, Trevor and Smith, 1940). An arrangement for transmitting television signals by frequency modulation, in which receivers designed for amplitude modulation receptions are used without change, has been demonstrated recently. Field tests with this form of transmission and reception are now in progress, and it is yet too early to evaluate its merit.

Another form of transmission, recently proposed, utilizes a combination of frequency modulation and amplitude modulation. In this system, the radiated carrier wave is amplitude modulated by those parts of the composite video signal corresponding to picture information, and is frequency modulated by the synchronizing pulse component. It is claimed that this type of transmission may be utilized by a receiver suitably designed for reception of single side-band, amplitude-modulated signals. This type of transmission is also undergoing field tests at this time, and its merits in practical operation are not yet fully known.

In order to accommodate the wide band of frequencies required in television transmission, and also to take advantage of desirable freedom from " sky-wave " transmission, television in recent years has resorted to the use of carrier frequencies above 40 Mc. As previously mentioned, carrier frequencies as high as 500 Mc. have come into use for television-relay purposes. In the United States eighteen channels, each 6 Mc. wide, have been set aside by the F.C.C. for allocation to television broadcasting stations. These channels are interspersed with other allocations in the region of the spectrum from 50 to 300 megacycles.

It is required that all signals for transmission of both video and audio intelligence fall within the 6 Mc. channel. Since all simple modulation processes for either frequency or amplitude modulation produce side-band components lying both above and below the unmodulated carrier in frequency, and since the minimum band-width of this double side-band spectrum, in any case, is equal

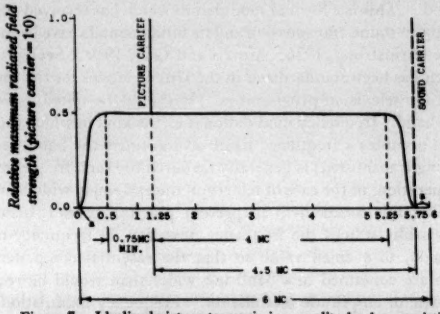


Figure 7. Idealized picture transmission amplitude characteristic.

a : Relative field strength of picture side band not to exceed 0·00005.

Drawing not to scale.

to twice the highest modulating frequency, it would seem that the 6 Mc. channel would not provide adequate transmission for television signals of the kind we have been considering, in which frequency components of 4 Mc. and more are present. Present practice in television broadcasting in the United States calls for the expedient of transmitting only one of the two side-bands generated for the higher video frequencies (Kallman and Spencer, 1940). It has been found that all of the information in the original signal may be reproduced at the receiver if only one side band is utilized (Poch and Epstein, 1937). In order to avoid distortions which occur with high percentages of modulation in single side-band transmission, present preferred practice calls for transmission of both side bands for the lower video frequencies, the attenuated side-band extending far enough to encompass those frequency components in the video signals which are of large amplitude in most cases. The standard amplitude characteristic of the signal

in a 6 Mc. channel as specified by the F.C.C. is shown in figure 7. The amplitude characteristic of a receiver designed to give proper response to such transmitted signals is shown in figure 8. It will be noticed that this receiver characteristic has a uniform slope in the vicinity of the carrier frequency and goes down to a 50 % response at the carrier frequency. It is this sloping characteristic in the receiver which enables it to translate frequency-modulated signals into amplitude-modulated signals as mentioned above in connection with two recently proposed systems of television transmission involving frequency modulation.

Although a video transmitter has all the general components possessed by a sound transmitter, there are several problems which are peculiar to wide-band video transmission (Blumlein, Brown, Davis and Green, 1938; Macnamara and Birkinshaw, 1938; D. G. F., 1939). The generation, amplification and radiation of ultra-short waves is a recent development in the radio art, but has been accomplished for services other than television, and the solution to the various problems involved are now fairly well known. The complications that arise in the case of television-transmitter design will not be gone into in detail here, but they

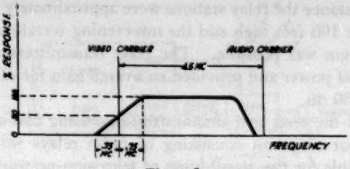


Figure 8.

include such matters as broad-band coupling circuits, special antenna (Lindenblad, 1939) and transmission-line design to provide uniform impedance and freedom from reflections over the broad band required, neutralization of high-power triode-amplifiers over a broad band, and design of suitable filters (Hollywood, 1939; Nergaard, 1939; Brown, G. H., 1941) for attenuating the unwanted side-band components without injuring the desired signal component.

The N.B.C. television antenna atop the Empire State Building in New York City is shown in figure 9.

The point-to-point transmission of video-signal intelligence has already been touched upon briefly. Because of the wide spectrum of the video signal there are very few facilities in existence which are capable of handling this type of signal. A nation-wide television service requires the distribution, to remote areas, of programme material originating in any one locality, as is now done by wire networks in sound broadcasting. When nation-wide sound broadcasting networks were desired, it was possible to utilize, in many instances, existing facilities which had been installed for the purposes of telephonic communication. It does not appear, however, that any existing facilities may be employed satisfactorily for television network purposes. Tests have been made, using both

wire line (Strieby and Weis, 1941) and radio circuits (Kroger, Trevor and Smith, 1940), which have shown that either medium is capable of providing the network facilities desired. Relative costs of the two media are not yet accurately known, but it seems likely that radio relaying will be adopted first in most cases because the facilities can be installed more quickly.

In planning a radio-relay system for television it is necessary to consider the proper antenna heights, antenna gain, spacing between relay stations and power radiated to give the most economical result. The propagation characteristics of the frequencies employed will have an important bearing upon the optimum balance of these various factors for lowest cost. The particular terrain in question will be a factor in determining the final layout, since use may be made of elevated points with lower tower structures, or greater distances permitted between stations.

A television radio-relay system, operating in the 500 Mc. region of the spectrum and employing frequency modulation, has been set up and demonstrated in the United States by R.C.A. Communications, Inc. (Kroger, Trevor and Smith, 1940). In this instance the relay stations were approximately fifteen miles apart, the antennas were 100 feet high and the intervening terrain was such that line-of-sight transmission was possible. The relay transmitters delivered approximately two watts of power and provided an overall gain for each repeater station of approximately 80 db.

The results of the tests and demonstrations, using this experimental relay set-up, showed that a system consisting of radio relays would be technically adequate and feasible for the distribution of television-network programmes.

(d) *Reception*

Home receiver. The signal received by the antenna preferably is conveyed to the radio receiver by some form of balanced line. This line may merely be a twisted pair, if the location of the antenna is such that a relatively short line can be used, and if the signal strength is high. On the other hand, if the feed line is long, and particularly if the field strength at the antenna is small, it may be necessary to use a two-wire, air-insulated transmission line or even a pair of coaxial cables.

The receiving equipment itself provides for the conversion of the radio-frequency signal into a video signal and for sufficient amplification for it to be used to operate the viewing device. It also separates the synchronizing impulses from the picture signal, employing the former to control the scanning pattern of the viewing device (Engstrom and Holmes, 1938, 39).

As with the conventional sound radio receiver, several alternative types of television receivers for home use are possible. The two most frequently used employ superheterodyne and tuned radio-frequency circuits.

A tuned radio-frequency receiver has five or six tuned stages of ultra-high frequency amplification, a detector, and several stages of video amplification.

Because of the difficulty of accurately tuning the many radio-frequency stages, it is rarely used when reception from more than one transmitter is desired (Brown, W. J., 1939, 40).

The superheterodyne receiver is made up of a radio-frequency element, a first detector and oscillator, an intermediate amplifier, and a second detector followed by a video amplifier. The radio-frequency element may merely be a broadly tuned circuit whose chief function is to match the antenna feeder to the first detector circuit, or it may be more complicated, containing, for example, a stage of radio-frequency amplification (Tyson, 1940). The radio frequency leaving this first stage is mixed with the output from the local oscillator and rectified by the first detector, giving rise to the intermediate frequency. The mixing is merely an addition of the incoming modulated signal and the locally generated unmodulated signal in a non-linear tube circuit which produces a beat effect between these two waves (Herold, 1940). The intermediate carrier and its side band are amplified by the intermediate amplifier, the other frequencies

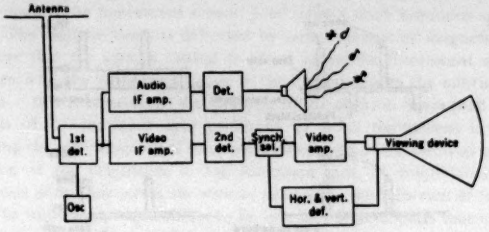


Figure 10.

present in the output of the detector being removed by the band-pass characteristics of the interstage coupling of this amplifier (Mountjoy, 1940). It has already been pointed out that the receiver-amplitude characteristic must have a sloping response in the vicinity of the carrier frequency if suitable reception of the transmitted signal, shown in figure 7, is to be had. This characteristic is obtained in the tuning of the intermediate, frequency-amplifier coupling circuits. Rectification of the mixed signals by a linear detector also leads to the intermediate carrier and side bands, but there are present, in addition, components resulting from the addition of various side-band frequencies. However, if the amplitude of the locally generated oscillation is large compared with the incoming signal, the amplitude of the unwanted combination of side-band frequencies will be small compared with those of the fundamentals. For practical reasons it is common to use a linear detector and a powerful local oscillator rather than a square-law detector.

After amplification, the intermediate carrier and side band are again rectified, this time by the second detector, which is also usually a linear detector. The

output of this element is a signal which corresponds exactly to the video signal generated by the pickup device, together with synchronizing impulses and any correcting signals inserted at the monitoring amplifier. A superheterodyne receiver is illustrated by the block diagram given in figure 10. As shown in this figure, the combination audio and video receiver employs a common first detector for the two signals, but separate *if* amplifiers and second detectors.

The level of the video signal leaving the final detector is, in general, not sufficient to actuate the viewing device. In consequence, it must be amplified by one or more stages of video amplification before it can be used. These stages are similar to those between the pickup equipment and the transmitter, being resistance-coupled and compensated to increase the high-frequency response (Foster and Rankine, 1941). Because of the nature of the coupling, the amplifier will pass only the a.c. components of the video signal. Therefore, it is usually necessary to provide some arrangement to re-insert the d.c. component, in order to establish the average illumination of the picture being transmitted.

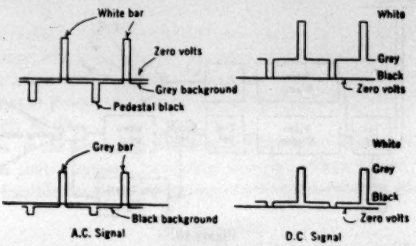


Figure 11.

By " d.c." signal is meant one for which the voltage relative to some fixed reference voltage corresponds to the absolute brightness of the point under the exploring spot at the transmitter. The diagrammatic oscillograms of figure 11 illustrate the distinction between the two types of signal. The d.c. level of the video signal can be re-inserted by making use of the pedestal or of the synchronizing impulse, because these bear a definite relation to either maximum dark or light of the picture.

A diode rectifier, or its equivalent, is employed for re-inserting the d.c. component to the signal at the terminals of the picture-reproducing device. This diode is similar in its action to a peak-reading vacuum-tube voltmeter, the d.c. voltage produced providing a bias for the picture reproducer which will maintain black level at the proper point on its characteristic (Foster and Rankine, 1941).

Some viewing device is necessary to convert the video signal into final visible reproduction, which is the ultimate objective of the system.

Two operations are necessary to perform the reconstruction of the image. The first consists of modulating a small but intense spot of light with the video signal; the second, of causing this spot to move in such a way that it sweeps out the scanning pattern on the viewing screen.

Broadly, there are two classes of viewing devices: electronic (Zworykin, 1929) and mechanical (Robinson, 1939). Of these, the former is the more widely used, particularly in home receivers. Very interesting results have been obtained with both systems as a means of producing large pictures for public viewing. The most widely used electronic reproducer is a modification of the cathode-ray oscilloscope tube known as the kinescope (Zworykin, 1929; Zworykin and Morton, 1940), the biggest change being the introduction of a control element to modulate the intensity of the cathode-ray beam. The kinescope consists essentially of an electron gun and a fluorescent screen enclosed in an evacuated container of suitable shape and size. The electron gun is a system of electrodes comprising an electron lens which drives a stream of sharply focused electrons against the fluorescent screen, producing a small luminous spot on the screen. The electron beam is deflected by either electric or magnetic fields in such a way that the spot is caused to move across the fluorescent surface in accordance with the scanning process in the camera where the television signal originates. At the same time, the intensity of the electron beam (and hence the brightness of the luminous spot produced by it on the screen) is controlled by applying the television signal to a control electrode of the electron gun. This modulation of the brightness of the luminous spot, in combination with its synchronous deflection across the screen, provides a reproduction of the original scene. An important requirement to be met in a kinescope is that modulation of the electron beam by variation of the potential on the control electrode must not result in appreciable variation of focus of the electron lens throughout the normal working range.

Home projection receiver. Very few of the television receivers now in use for home reception employ optical magnification of the received image. The cathode-ray screen is viewed either directly or by means of a mirror located in the lid of the receiver cabinet. Although a few types employ picture tubes having a 5-inch or 9-inch screen diameter, the most popular size is a 12- or 14-inch cathode-ray tube giving an image approximately 8×10 inches. A few receivers employing a 20-inch cathode-ray tube have been demonstrated (Goldsmith, 1940), but these very large cathode-ray tubes have not come into general use.

While the present receivers utilizing the 12-inch cathode-ray tube have given very satisfactory results from the standpoint of both picture resolution and brightness, there has been a general desire for larger television pictures, of the order of 18-inch to 24-inch in width, with increased contrast range, comparable to home motion pictures. It appears that the best engineering solution to the problem of providing such pictures is to produce a very intensely illuminated image on a small cathode-ray tube (Zworykin and Painter, 1937) and project it,

optically magnified, on a larger viewing screen. Such an arrangement has many advantages, being more compact and economical than a very large cathode-ray tube directly viewed, and the indications are that some designs of projection receivers could be built in quantity to sell at a very reasonable price.

Early in 1941, one manufacturer demonstrated samples of a home-projection receiver having several novel features. This utilized a special 5-inch projection cathode-ray tube operated with a second anode voltage of about 27 kv. This voltage, as well as the first anode voltage, was generated by means of triode oscillators operating at a low radio frequency, the R.F. voltage being stepped up, by means of an auto-transformer arrangement tuned to resonate at the frequency of the oscillator, and rectified. Focusing was accomplished by tuning the oscillator supplying the first anode potential. It was pointed out that, although a power interlock system was provided to prevent the user from coming in contact with the high voltage, due to the smaller amount of energy stored in the R.F. high-voltage filter system, this power supply does not have the same lethal properties as many conventional power supplies of much lower voltage, but greater stored energy, as used in present direct-viewing television receivers.

The television image was projected via a suitable field flattener and a lens having a focal length of 5 inches and a rated aperture opening of f/2. The effective speed of the lens and the contrast range of the projected image were enhanced by coating the glass surfaces to prevent internal reflections. The television image was projected from the rear of a translucent viewing screen giving a picture 13·5 inches by 18 inches in size with a peak brightness of 10 foot-lamberts and a picture resolution of over 400 lines. The whole assembly was designed to fit into the same cabinet as a standard commercial 12-inch direct-viewing receiver.

Theatre projection. During the past several years, projection television equipment designed for theatre use has been demonstrated from time to time. Both the Baird system, using a projection cathode-ray tube operated at high voltages and a fast projection lens, and the Scophony system, utilizing a supersonic light valve in combination with mechanical scanning, have given very creditable results. The most recent development in this field is the newly designed theatre projection equipment (Zworykin and Painter, 1937; Maloff and Tolson, 1941; Law, 1937) demonstrated by the R.C.A. Manufacturing Company in a New York theatre early in 1941. The equipment utilizes a 7-inch white screen cathode-ray tube operated at an anode voltage of 70 kv. and producing a maximum luminous flux output of 1200 lumens. This light is collected by an optical system consisting of a spherical mirror and an aspherical lens or correction plate, which combine to form a projection system free from spherical aberration or coma. With a mirror 30 inches in diameter, the projection system has an efficiency of 25 %, delivering a luminous flux of 300 lumens to a viewing screen of full motion-picture size (15 × 20 feet). An idea of the increased efficiency of

this special optical system can be gained from the fact that a conventional arrangement using a lens with an f/2 aperture would have an efficiency of only 3¾ %. The picture definition obtainable with the equipment was over 400 lines, when operated at the scanning rate of 441 lines, 60 fields, 30 frames per second. With a projection screen having a 2:1 directivity, the picture had an effective peak brightness of over 2 foot-lamberts, which is comparable to the range of 5 to 20 foot-lamberts peak brightness obtained with film projection on the same size screen in the better motion-picture theatres. A screen with a 5:1 directivity gave a peak brightness of 5 foot-lamberts with the projection television system.

§ 3. SPECIAL PROBLEMS

(a) *Pickup tubes*

The Iconoscope. The basic limitation of mechanical scanning devices is so well known and so obvious that it will not be commented upon here. Practically all modern television pickup devices employ some type of electronic scanning. The one which has enjoyed most extensive use is the iconoscope (Zworykin, Morton and Flory, 1937; Janes and Hickok, 1939; McIlwain, 1939; Fink, 1940; Zworykin and Morton, 1940).

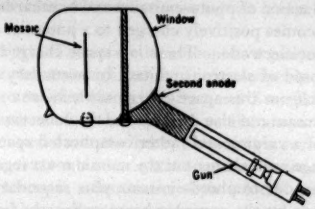


Figure 12.

The iconoscope consists of a photo-sensitive mosaic and an electron gun assembled in a highly evacuated glass envelope. The gun is an electron-optical system serving to produce a fine pencil of cathode rays which is made to scan the sensitized side of the mosaic by means of a suitable magnetic, or electrostatic, deflecting system. The mosaic in the normal iconoscope is a very thin mica sheet covered on one side with a vast number of minute silver globules, photo-sensitized and insulated from one another, and coated on the other with a metal film known as the signal plate. This metal film is coupled, on the one hand, to the silver elements by capacity through the mica and, on the other, to the video amplifier through a signal lead sealed into the bulb. The drawing of figure 12 shows a section through a typical iconoscope. It will be seen that the electron gun is not normal to the mosaic surface, but is displaced from normal by an angle of approximately 30° in order to provide clearance for the optical system employed to focus the light image upon the mosaic.

The optical image is projected on to the silver-coated side of the iconoscope mosaic. Each element, being photo-emissive, accumulates charge by emitting photo-electrons. Thus, the information contained in the optical image is stored on the mosaic in the form of a charge image. The scanning beam, sweeping across the mosaic in a series of parallel lines, releases the charge from each element in turn and brings it to equilibrium, ready to start charging again. The change in charge in each element induces a similar charge in the signal plate and, consequently, a current pulse in the signal lead. The train of electrical impulses so generated constitutes the picture signal. The iconoscope differs from all previously developed pickup devices in being able to store information for each element of the image during the entire scanning cycle, thus giving tremendous increases in sensitivity over non-storage devices.

The simple theory of operation of the iconoscope given above is complicated by the secondary emission of electrons which occur when the high-velocity scanning beam strikes the mosaic. Due to the fact of this secondary emission, the scanning beam, instead of merely replacing electrons emitted due to photo-emission, actually drives the scanned area to a more positive potential. In fact, the element under the scanning beam is the most positive on the entire mosaic. Under the combined action of photo-emission and bombardment by the scanning beam, the mosaic becomes positively charged to a potential very nearly the same as that of the collector electrode. There is a space charge in the region in front of the mosaic composed of electrons emitted by secondary emission and photo-emission. Electrons from this space charge settle more or less uniformly over all elements of the mosaic and also drift toward the collector electrode, depending upon the influence of a varying and rather complicated space distribution of the fields. On the average each element of the mosaic must regain the same number of electrons as it loses, due to photo-emission plus secondary emission. At the point of bombardment by the scanning beam, when the beam current is fairly high, the surface of the mosaic assumes a voltage between plus 2 and plus 3 volts with respect to the collector. The most negative area of the mosaic lies always just ahead of the scanning beam, and its potential is usually minus 1½ volts. For lower beam currents the difference in potential between the most positive and most negative points on the mosaic is somewhat less.

The mechanism of the generation of the picture signal may now be considered on the basis of the above factors. Fundamentally, it is due to the fact that the potential of an illuminated element of area, just before being reached by the scanning beam, is more positive than one which has not been illuminated. Since the mosaic is driven to the same positive equilibrium regardless of its initial potential, charge will be released at a different rate when the beam is traversing the lighted region from that when scanning an unlighted area. The change in current reaching the collector constitutes the picture signal.

The Image Dissector Tube. The image dissector tube (Farnsworth, 1934; McIlwain, 1939) is an electronic television pickup device in which the light

image is focused upon a photo-cathode surface and the electrons emitted therefrom in accordance with the light values at each point are brought to focus, by means of an electronic lens system, in the plane of a small aperture which is of picture element size. The electrons emitted from the photo-cathode are deflected both horizontally and vertically by means of deflecting coils, so that the electron image in the plane of the fixed scanning aperture is swept across the aperture in order that electrons from the various elementary areas on the photo-cathode may contribute their charge in proper turn to the collector inside the scanning aperture. An electron multiplier is associated with the collector inside the scanning aperture so that the minute currents collected may be intensified before being impressed upon the conventional video amplifier.

The important difference between the image dissector tube and the iconoscope is the lack of any storage of picture information in the former. As a result of this, its sensitivity is considerably lower. On the other hand, it is free from spurious signals which are produced by the complications arising from secondary emission in the iconoscope. In recent years the image dissector tube has found its most important application in the televising of film, where its relatively low sensitivity may be overcome by the use of bright light in projection and where its freedom from spurious signals is especially advantageous (Jensen, 1941).

The Image Iconoscope. The image iconoscope is a pickup device consisting essentially of the same elements as a standard iconoscope, but with the addition of an electronic image amplifier (Iams, Morton and Zworykin, 1939). In this tube the optical image is focused upon a semi-transparent photo-cathode surface. Electrons emitted from the other side of this surface are brought to focus, by means of an electronic lens system, upon a mosaic surface similar to that in a standard iconoscope. Because of their velocity, caused by a suitable acceleration potential, these electrons produce a charge upon the mosaic by secondary emission which corresponds to the original image, and this charge is then scanned and a video signal generated in the same manner as in the standard iconoscope. The addition of the electronic image amplifier gives this tube a greater sensitivity than the standard iconoscope. Its English counterpart, the super-emitron, came into use in some of the regular television operations of the B.B.C. during the last two years of service prior to the war.

The Orthicon. During the last two years of television programme operation in the United States a new type of iconoscope, known as the orthicon, has come into use in those applications where high sensitivity is of paramount importance (Rose and Iams, 1939 a, 39 b). This tube also makes use of the storage principle, and may be said to differ fundamentally from a standard iconoscope in the employment of a low-velocity cathode-ray scanning beam. The importance of this difference lies in the fact that the low-velocity electrons do not produce any secondary emission upon striking the mosaic, and hence do not produce the space-charge effects and attendant spurious signals which are present in the

standard iconoscope and image iconoscope. This tube makes use of a mosaic similar to that employed in the iconoscope except that the signal plate is semi-transparent and the optical image is focused upon the mosaic by transmission through the signal plate. Although this arrangement results in a reduction of sensitivity due to imperfect light transmission through the signal plate, it allows the light optical system and electron optical system to be on opposite sides of the mosaic, and the axis of the electron gun, therefore, can be at right angles to the mosaic, whereas in the iconoscope it must be off the normal by approximately 30° in order to avoid physical interference with the light optical system.

The low-velocity scanning beam employed in the orthicon creates special problems in deflection and in focusing. In the type most commonly used, focusing of the scanning beam is by means of an axial magnetic field, and vertical deflection is electro-magnetic. Horizontal deflection is provided by means of an electrostatic field operating in conjunction with the magnetic focusing field, the electrostatic plates in this case being horizontal, or parallel to the direction of deflection.

In its present state of development the orthicon tube is not capable of as high resolution as the best standard iconoscope, due to imperfect focusing of the low-velocity electrons in the scanning beam at the surface of the mosaic. On the other hand, its greater sensitivity, due to a combination of better storage and freedom from spurious signals, has opened new possibilities in television programme operation, and its resolution is adequate for many purposes.

Future developments. Although the pickup tubes available at present for television operation have sufficient sensitivity to give good pictures with reasonable light levels, it would, nevertheless, be desirable to increase the sensitivity. The orthicon tube represents an improvement over the standard type of iconoscope in this respect, and also in its freedom from dark-spot signal. Even the orthicon, however, does not give the television camera the sensitivity necessary to equal the performance possible in photographic work when using the recently developed super-sensitive film emulsions. Many schemes have been suggested for improving the sensitivity of the television pickup devices and a considerable amount of experimental work has been done and is still in progress on some of these schemes. No attempt will be made here to cover the entire field, but some mention might be made of two of the most promising proposals.

The principle of the electron multiplier, which has made possible such substantial improvements in the sensitivity of photo-electric tubes, has also been applied to television pickup tubes. The image iconoscope, previously described, embodies one stage of electron multiplication. Experimental tubes have also been made in which a portion of the electrons emitted from the iconoscope mosaic under the impact of the scanning beam is led into a series of electron multiplier stages. A several-fold increase in sensitivity, compared with standard iconoscopes, has been obtained by this method, but some difficulty has been experienced in obtaining uniform pickup by this method of the electrons from

all parts of the mosaic. A more promising application of this principle is the use of electron-multiplier stages in conjunction with the low-velocity scanning-beam type of tube (orthicon). In this tube those electrons from the scanning beam which are not absorbed in equalizing the charge on the mosaic produced by photo-emission are turned away from the mosaic and are collected in an electron multiplier. In this case there are no complicating factors due to secondary emission (until the electron-multiplier stages are reached), and this tube shows considerable promise of a several-fold increase in sensitivity over the standard type of orthicon, which is itself several times more sensitive than the standard iconoscope. Tubes have been built in accordance with this principle and operated in the laboratory, but the design is not yet ready for tests in the field.

Since spurious dark-spot signal inherent in the standard type of iconoscope tube is eliminated by use of a low-velocity scanning beam, it now appears likely that future developments in the field of pickup tubes will all utilize the low-velocity scanning beam. Another interesting development in this field is that of a tube in which an image multiplier stage is employed ahead of the mosaic just as in the case of the image iconoscope, but with a low-velocity scanning beam for the mosaic.

(b) *Colour television*

A study of the history of the motion-picture industry reveals that experimenters and inventors began to attempt to add colour to the monochrome reproductions almost as soon as the first crude pictures were produced. The reproduction of scenes in natural colour in most cases adds a desirable quality and, provided there are no serious sacrifices in other picture characteristics, it is decidedly worth while. It is only within recent years, however, that colour-motion pictures have achieved the technical perfection which makes possible their wide-scale acceptance, and even to-day many important feature films are made in monochrome.

In television, also, the desire to add natural colour to the reproduced television picture has stimulated the minds of inventors and experimenters almost from the beginning of the television art. It was not until the 1920's that colour experiments were demonstrated that seem worthy of mention. The literature of that decade describes several different systems for colour television, prominent among which are reports on the work of the Bell Telephone Laboratories (Ives and Johnsrud, 1930; Ives, 1931) in the U.S., and J. L. Baird (Tiltman, 1928) in England. In 1925 a proposal for an all-electronic colour television system was made by V. K. Zworykin (1928).

During the past ten years a number of systems for the transmission of television images in natural colours have been described in the literature. Some of these, while theoretically sound, have for practical reasons never advanced beyond the proposal stage, but others have actually been demonstrated with varying degrees of success. The same general principles of colour analysis and

synthesis, long used in colour photography (Luckeish, 1927), have been utilized. According to these principles, the colours in the original image are split up into relatively narrow spectral bands corresponding to the three primary colours, and these are transmitted separately. At the receiving end, the three video signals so generated actuate light sources of the corresponding colours to reproduce the proper shade of colour for each elementary picture area. While such a procedure would seem to involve a compromise in colour reproduction, inasmuch as a continuous spectrum cannot be transmitted accurately, it is a well-known fact—repeatedly demonstrated by three-colour printing and three-colour photographic processes—that as far as the eye is concerned a highly satisfactory result can be obtained. The reason for this is two-fold: the inability of the eye to analyse colours accurately, and the fact that monochromatic colours occur only very rarely in nature or in the scenes to be reproduced.

As in photography, a three-colour system for television may use either the additive method, with red, green and blue as the primary colours; or the subtractive method, in which the primaries are yellow, blue-green and magenta. An example of a system using the subtractive method is a recent proposal involving the use of special cathode-ray tubes which operate as light valves for picture reproduction (Rosenthal, 1940a, 40b). The cathode-ray target in these tubes consists of a crystal which is normally transparent, but which will absorb light in a narrow wave-band when bombarded by electrons, the effect persisting for a fraction of a second after cessation of bombardment. In a reproducing system for a three-colour subtractive process, the light would pass through three such tubes in tandem, each one providing absorption of a different primary colour component from the original white light in accordance with signal modulation of its electron scanning beam. While this proposal is novel and interesting, there is no report of any such system having been set up and tested.

Inasmuch as most of the colour-television systems which have been set up for experiment and demonstration have employed the additive process, the subtractive process will not here be dealt with further.

Various two-colour systems using blue-green and orange-red filters for analysis and synthesis have been proposed, and one of these has been demonstrated recently (Bartlett and Caldwell, 1941). The fidelity of colour reproduction obtainable is in general inferior to that with three-colour systems, but the process has some merit since, in comparison with three-colour processes, it results in simplification of equipment and permits the transmission of more detail within a frequency band of fixed width.

In addition to the two-colour and three-colour additive and subtractive classifications already referred to, colour-television systems may be subdivided further under the general heading of *cyclic systems* and *simultaneous systems*, according to the manner in which the electrical signals corresponding to the primary colours are transmitted and received.

A cyclic system is one in which the video signals corresponding to the three

primary colours are transmitted and received on a time-division basis over a single communication channel. An example of a cyclic system is one in which complete scanning fields of the primary colours are transmitted according to a pre-determined sequence. Since three fields are required to complete the red, blue, green colour cycle, it is evident that with this system the avoidance of flicker requires that the vertical deflection rate be substantially increased above that normally used for monochrome of comparable definition.

A typical simultaneous system employs three pickup tubes, one for each of the primary colours, continuously focused on the scene, three separate transmission channels for the electrical signals, and optical means for recombining the light from three reproducer tubes into a single image at the receiving end. In such a system the proper colour-selecting filters are placed in front of each pickup device and in front of its corresponding reproducer at the receiving end (or each reproducing kinescope may have a screen phosphor which produces, directly, light of the proper primary colours). Parallax can be avoided by using a single lens at the sending end with a suitable optical arrangement for focusing three identical images on the three pickup tubes with their colour-separation filters. Such optical combinations are extensively used in cameras for three-colour photography and will not be described in detail here. The simultaneous system has several advantages: (i) The vertical deflection rate need not be greater than in a comparable monochrome system to avoid flicker; (ii) the apparent brightness of the received image will generally be greater than with a cyclic system, since the receiver screen is continuously illuminated with all three primary colours; and (iii) there are no mechanical moving parts.

The main practical disadvantage encountered in applying this method is the difficulty in obtaining sufficiently accurate deflection of the cathode-ray beams at both sending and receiving ends, so that satisfactory registration of the three images may be obtained. It is interesting to note here the striking similarity to the problem that existed for some years in the field of colour cinematography with respect to accurate registration of the coloured images transferred from the several " taking " films to a single coloured film for projection. It is hoped that the problem of registration in three-colour simultaneous television may be solved by the same type of painstaking work that finally produced satisfactory colour-motion picture films.

In passing, it might be mentioned that a three-channel simultaneous system does not necessarily involve the use of three separate transmitters and receivers. It is possible to modulate a single carrier by three subcarriers to produce a signal capable of being sent out by a single transmitter. The resulting single carrier and side bands can be demodulated in a suitably designed receiver to separate the signal into the three original channels.

There are numerous possible variations of the simultaneous and cyclic systems as outlined here, and no attempt will be made to survey all of the many systems that have been proposed.

Recently there has been a revival of interest in a cyclic method of colour television which bears some similarity to a cyclic film system (Kinemacolour) that enjoyed favour for a brief span in the early days of colour-motion pictures (Luckeish, 1927). The arrangement in this type of system provides for the use of electronic pickup and reproducing devices in connection with colour filters moved synchronously by mechanical means in front of the camera and receiver tubes. While it might seem to many that it is a backward step to incorporate mechanical rotating machinery in a television system which is otherwise completely electronic, the method is attractive from the standpoint that it produces better results than any other immediately available. Experiments, using semi-mechanical cyclic systems, are now in progress at several laboratories and, although the histories of both colour-motion pictures and television itself tell us that such systems will not survive the advent of a practical all-electronic colour television system, it is undoubtedly true that much is being learned through these experiments which will be useful and valuable in the future art, no matter what system may be the prevailing one ultimately.

During the latter part of 1940, considerable interest was shown in the results obtained in demonstrations by the Columbia Broadcasting System of a cyclic system (D. G. F., 1940; Goldmark, 1940) using 60 frames per second, 343 lines interlaced 2 to 1. Electronic pickup and cathode-ray reproducing tubes were used in connection with mechanically rotated filter wheels fitted with red, green and blue filters. Complete colour fields of the primary colours were transmitted at the rate of 120 per second, making the colour-flicker rate 40 per second and giving a colour-scanning coincidence rate of 20 per second. (The colour-scanning coincidence rate is the rate at which each scanning line goes through the complete cycle of three colours.) The colour-scanning coincidence rate was one-half the colour-flicker rate by virtue of the 2-to-1 interlace used. With these standards, and with a given communication channel, a loss in total resolution of approximately two to one, compared to conventional monochrome systems, is accepted to allow for the addition of colour. Also, visible flicker is present in the reproduced picture under some conditions.

A three-phase amplifier section with separate gain control on each of the phases permitted independent variation of the signal corresponding to each of the three primary colours, thus permitting convenient adjustment of colour balance in the reproduced picture.

While the recent tests and demonstrations by C.B.S. and others have served to show that excellent colour reproduction can be obtained with such a cyclic system, there is considerable doubt whether the sacrifice in resolution will be acceptable except in certain types of scenes. When attempting to add colour to an existing monochrome system with a fixed communication channel, it must be realized that some compromise is necessary. This follows when it is understood that the addition of colour requires the transmission of a greater amount of information in a given time than for an equivalent monochrome picture. In

monochrome transmission it is necessary only to convey information as to the position and illumination intensity of each picture element, but when colour is added, we must not only transmit information about the position of each picture element, but also the individual intensities of the red, green and blue components which combine to reproduce its colour.

Since a compromise must be made with respect to resolution and/or flicker rate when applying the cyclic method to a system applicable to the video band width available at present, it is of interest to consider the possible choices. Some of them are:

(*a*) The present scanning rates of 525 lines per frame, 30 frames per second can be retained and colour added by transmitting sequential colour fields. Under these conditions, the colour-flicker rate will be 20 per second, and the colour-scanning coincidence rate will be 10 per second. Although this combination produces the same horizontal and vertical picture resolution as for monochrome conditions, actual tests show that the flicker produced is intolerable, as would be expected.

(*b*) Keeping the same number of horizontal scanning lines per frame, and the same 2 to 1 interlace ratio, the field rate can be doubled to 120 per second. In order to obtain the same horizontal resolution as before, however, the video band width must be doubled. If we keep the same band width (4·25 Mc.), the horizontal resolution will be halved, but the flicker will be greatly reduced.

(*c*) We may keep the same number of lines per frame, 525, the same number of frames per second, 30, but increase the field frequency and the interlace ratio. This arrangement gives full horizontal and vertical resolution, and cuts down field flicker. However, experience has shown that any interlace ratio greater than 2 to 1 is unsatisfactory because the eye can partially resolve the individual scanning lines and an objectionable " creeping " effect of the lines is seen in the picture.

(*d*) An alternative to (*b*) above would be to use double the vertical scanning frequency and double the frame rate, but reduce the number of horizontal picture scanning lines until horizontal and vertical picture resolution are approximately equal.

Total resolution would, however, still be limited as in (*b*). It has recently been reported (Baldwin, 1940) that the distribution of definition, between the horizontal and vertical dimensions, is not critical for a given overall resolution.

The present 6·0 Mc. channels allocated in the United States for broadcasting television and associated sound make available an effective video-frequency band width of approximately 4·25 Mc. at most. With some types of scene, the lower image definition obtainable when applying the three-colour cyclic system to this limited band width in place of the standard monochrome system is compensated for by the colour resolution obtained. However, since in general there are various psychological and other complex factors involved, comprehensive field tests will be required to determine the usefulness of such a colour system in the limited band width.

A comprehensive field test of colour television was started in March 1941 by the National Broadcasting Company and is still in progress as this *Report* is written. The system tested has used 441 and 343 lines per frame, and 120 fields, 60 frames per second, interlaced 2 to 1. Consecutive scanning fields are transmitted in the primary colours in a red, blue, green sequence. The colour television signals, comprising both film and direct pickup subjects, originating in the N.B.C. studios have been broadcast over Station W2XBS, the television transmitter normally used for monochrome broadcasts in New York. These signals have been observed at various points in the Metropolitan area, and at distances up to 50 miles, by means of colour receivers especially prepared for the field test.

An orthicon (Rose and Iams, 1939 b) camera equipped with a specially designed colour wheel and fitted with standard Wratten no. 25, no. 58 and no. 47 red, green and blue filter transparencies has been used at the sending end for both film and direct pickup subjects. Approximate colour balance was obtained by using either high-intensity carbon arcs or fluorescent lamps having a colour temperature of 4000–5000° K. for illuminating the scene, and by adjusting the effective transmission of the sending-end colour filters. The camera tubes used have essentially uniform response for the three primary colours, so that the remaining small colour un-balance can be readily corrected by means of individual electrical colour-gain controls operating in a three-phase amplifier provided for that purpose.

The receivers utilize kinescopes having a short-persistence white sulphide screen-material, giving substantially uniform colour response over the visible spectrum when viewed through a rotating wheel containing no. 25, no. 58 and no. 47 Wratten tricolour filters. Most of the receivers used so far are of the standard 9-inch viewing type, modified to accommodate the colour wheels and driving motor. Additional power-supply filtering had to be added because of the difficulty in obtaining good interlace when operating at 120-cycle field frequency, which is double the a.c. power-supply frequency. A projection type colour receiver giving an image 18 × 13 inches and $\frac{1}{2}$ inch in size was also tested.

This field test is still in progress, but the following tentative results have been obtained:—

(*a*) Those who have viewed the transmissions have been well pleased with the fidelity of colour rendition, and some expressed the opinion that the colours reproduced were equal to the best in coloured motion-picture films.

(*b*) Many observers have been disappointed by the low value of resolution possible with the system. This is not objectionable in close-up scenes, but is very apparent in average scenes.

(*c*) The obtaining of good interlace on the receivers operating at 120 fields turned out to be a major problem, inasmuch as a very small amount of residual 60-cycle hum in the deflection, video or high-voltage supply circuits completely destroys interlace. If present in the horizontal deflection circuit, 60-cycle hum

cuts down picture resolution. Similar problems were encountered in the pickup equipment associated with the camera.

(*d*) In the field tests, some flicker was observed, mostly on white high-lights in the picture. Another effect directly related to the vertical deflection rate, and therefore to the colour-repetition rate, is that of " colour break-up ". This takes place when objects move rapidly in the picture field. A white handkerchief moved rapidly will appear as a number of red, green and blue blotches. Also an observer watching a coloured image, on suddenly turning away from the picture, or upon blinking his eyes, will momentarily see a completely red, blue or green screen, depending on the timing of his movements. On ordinary gestures no serious colour break-up was observed.

Pickup tube problems in colour television. The cyclic method of colour-television transmission places additional burdens upon the television-pickup tube employed in the camera. The most obvious of the new problems relates to the colour response of the tube. The spectral response of the pickup tube will, of course, have a direct effect upon the fidelity of the reproduced colour. While it is true that the overall colour fidelity is also affected by other characteristics of the system, including the spectrum of the light used to illuminate the scene being televised, and while these variables might be manipulated to compensate for variations in the colour response of the pickup tube, experience so far indicates that the most desirable colour-response characteristic for the pickup device is essentially a uniform one over the entire visible spectrum. A good approximation to such a response is possible in pickup tubes of the orthicon or iconoscope type, but suitable uniformity in all tubes on a production basis is not easy to achieve at the present time.

The effective sensitivity of a pickup camera in the three-colour cyclic process is substantially less with the same lens than the same camera would have with black and white operation. The exact ratio between these two sensitivities is not easy to calculate or measure, but it is in the order of four or five to one in favour of black and white operation. There is a three to one loss in sensitivity inherent in the cyclic process and some additional loss due to imperfect transmission through the three-colour filters. This loss in sensitivity means that in the televising of a given scene by the cyclic-colour method, a greater amount of light must be used on the scene, a faster lens employed on the camera, or a more sensitive pickup tube used, than when televising the same scene in black and white. (We have ignored the difference in sensitivity between colour and black and white, which is due to the difference in scanning standards usually employed, since it is believed to be smaller than the other factors mentioned.)

It has already been shown that modern high-definition television is possible only with electronic pickup devices employing the storage principle. Because of the added sensitivity required in the cyclic colour process, the importance of these storage devices for direct pickup is still further emphasized. However,

with the interlaced scanning and colour sequence cycle proposed for colour television, the storage property in the pickup tube results in one undesirable effect. When light from the scene being televised passes through any one of the three filters in the optical system, the light image corresponding to that colour is focused upon the entire mosaic in the pickup tube. This creates an electrical charge on all elements of the mosaic corresponding to the light intensity of this colour component of the image at each point. Since, with the interlaced scanning generally favoured, only alternate lines in the picture will be scanned before the colour filter is replaced by one of a different colour, it follows that the unscanned lines on the mosaic will retain an electrical charge corresponding to the preceding colour image, upon which the electrical charge of the new colour image will be superimposed. This results in colour mixture and impaired colour fidelity in the reproduced image. Actually, with present pickup tubes the colour mixture is not as serious as might at first be supposed because of imperfect focusing of the electron beam which scans the mosaic. Since some of the electrons in the beam fall outside the nominal scanning-spot area, the effective width of the scanning spot is greater than one scanning line, and a partial equalization of the charge on the unscanned lines results. This imperfection in present-day pickup tubes has made possible the relatively good colour fidelity demonstrated in three-colour cyclic television transmissions, but it causes a corresponding reduction in resolution.

Several schemes have been proposed to eliminate this difficulty caused by the charge remaining on the unscanned lines at the end of each field scanning. One of these proposals (D. G. F., 1940; Goldmark, 1940) involves the shifting (vertically) of the optical image by a distance equal to the height of one scanning line at the end of each field scanning and a corresponding scanning control which causes the electron beams always to scan the same lines on every field. In this way the charge on the unscanned lines presumably would be ignored. Other proposals for solving this problem include various means of equalizing the electrical charge on the unscanned lines by auxiliary sources of electrons or by causing the main electron beam to sweep across these lines upon its return during the horizontal or vertical blanking interval. There has been no report of a demonstration of any of these in operation.

Either the iconoscope or the orthicon type of pickup tube may be used in cyclic colour television, and both have been used. However, the orthicon type of tube is preferred for two reasons. First of all it is substantially more sensitive than the iconoscope type and, as previously emphasized, sensitivity is quite important. In the second place, the iconoscope tube has a spurious output (dark spot signal) which is even more troublesome with colour than in black and white operation. Since the dark spot signal is not constant, but varies with the amount of light and type of scene, it follows that the dark spot signal is not always the same for the three colours in a cyclic colour system. It is possible

to provide separate compensation for the dark spot signal for each of the three colours, but this is rather complicated and, therefore, undesirable. Another advantage of the orthicon tube for colour work is its linear characteristic as regards signal output versus light input.

It has been found experimentally that colour mixture is substantially reduced in the orthicon type of tube by increasing the thickness of the dielectric between the mosaic and signal plate. Special orthicon tubes having this lower capacitance between mosaic and signal plate were employed in the recent colour-television field tests and demonstrations previously reported.

Present status and future development of colour television. In recognition of the present interest in colour television, the Federal Communications Commission in the United States recently announced preferred values for standards for experimental colour transmissions. These tentative preferred standards fix a field frequency of 120, a frame frequency of 60, and 375 lines per frame interlaced 2 to 1. The preferred colour sequence was given as red-blue-green. The Commission has asked that television station licencees carry out tests of colour-television systems.

(c) *Scanning and synchronizing*

Consideration of number of lines. The problem of the transmission of a picture of finite detail and discontinuous or limited motion has no unique solution (Zworykin and Morton, 1940). Of the several possible methods, that involving the process of scanning has been chosen as the most practical for use in all important systems providing a service or undergoing experimental development at the present time. This scanning process consists of moving an exploring element or spot over the image to be transmitted in a periodically repeated path covering the image area. The exploring element is so arranged that it will generate a signal which indicates the average brightness of its instantaneous position. This signal is transmitted over the communication channel to the reproducing spot, the brightness of which is controlled by the signal. The reproducing spot is caused to move over the viewing screen in a path similar to, and synchronous with, the path of the exploring element. Thus the reproducing spot reconstructs at the viewing screen, both in magnitude and distribution, the brightness of the image area.

A variety of configurations has been proposed for the scanning path covering the image area and the viewing screen. These include spiral scanning, sinusoidal scanning, and so forth. In what follows, we will confine our discussion to straight-line scanning, since that has been generally adopted as the most desirable form.

In order to reproduce exactly the scene being televised, there must be exact synchronism of the scanning patterns in the pickup and reproducing devices and there must be a one-to-one correspondence between the brightness of the

reproducing spot and the brightness of the scene at the point corresponding to the position of the exploring element.*

In order to maintain the illusion of continuous motion, the separate pictures showing the different steps of the motion must follow one another rapidly. The motion-picture industry has adopted twenty-four frames per second as a standard, partly to allow for better sound fidelity in the usual sound-on-film arrangements than could be had at a lower frame-repetition rate. This rate, higher than absolutely necessary, insures smooth motion for most types of pictures. In television it has been found desirable that the frame-repetition rate bear a simple relation to the frequency of the commercial a.c. power supply (Zworykin and Morton, 1940). This minimizes undesirable effects of residual hum from rectifiers and stray fields, and is also helpful in practical use of motion-picture-film projectors with the television system which are of the intermittent type. Thus the frame-repetition rate chosen for use in the United States is thirty per second, the power-supply frequency most widely used being 60 cycles per second, while in England and on the European continent the frame-repetition rate most widely used is 25 per second, due to the prevalence of 50-cycle a.c. power supply. Actually, the rates of 25 or 30 frames per second, while fulfilling most requirements of continuity of motion, are not high enough to avoid flicker effects. To meet this situation, interlaced scanning has, in recent years, been almost universally adopted for television. In this special form of straight-line scanning, the spot, instead of moving across the horizontal lines in simple sequence, scans the odd-numbered lines first, then the even-numbered lines. It thus covers the field from top to bottom twice in order to scan completely all portions of the picture and, therefore, the field frequency is twice the frame frequency.

It is apparent that this form of scanning is not restricted to the simple interlacing mentioned. The scanning may be interlaced in accordance with much more complicated schedules, requiring three or more field excursions completely to scan the picture. The simple two-to-one interlace process has, however, been most widely adopted, since experience has shown that with this arrangement the flicker rate for most types of scenes is equal to the field-repetition rate, while for the more complex forms of interlacing, troublesome effects such as interline flicker, and an appearance of crawling or weaving, begin to show up.

One of the most controversial subjects in the field of television in recent years relates to the obtaining of optimum resolution with a given arbitrary limitation on the band width available in the communication channel (Wheeler and Loughren, 1938; Kell, Bedford and Fredenhall, 1940; Goldmark and Dyer, 1940; Baldwin, 1940). It has been generally assumed that the reproduced

* In actual practice such a linear relationship in brightness may not be possible or even desirable. In many scenes which it might be of interest to transmit, the range of brightness is far beyond the capabilities of many of the reproducing systems. As in photographic practice, it has been found both expedient and desirable to utilize non-linear relationships in this characteristic, the preferred relationships being governed by many of the same considerations as apply in photographic technique (Maloff, 1939 ; Fink, 1941).

picture should have approximately equal horizontal and vertical definition. Recent study (Baldwin, 1940) of the problem, however, by means of simulated television images has shown that the optimum distribution between horizontal and vertical resolution with a given fixed amount of total definition is not at all critical. It was, in fact, found that to produce a definitely discernible deterioration in overall quality a change of approximately two to one in the distribution between horizontal and vertical definition was necessary when the total definition remained constant. This means thatin a television system having a communication channel of fixed-frequency band width the optimum number of scanning lines is not critical within a two-to-one range. Other factors, of course, such as appearance of the scanning-line pattern, enter into a determination of a preferred number of scanning lines. In England the public television service in operation prior to the outbreak of the war employed a 405-line standard (Blumlein, Brown, Davis and Green, 1938), and in Germany 441 lines was the preferred value in 1939. In the United States, the standards recently set by the Government for commercial television operation call for 525 lines.

Vertical resolution in a television picture is governed primarily by the number of scanning lines. Although vertical resolution may be increased with a fixed number of scanning lines by the use of very small scanning apertures at the transmitter and reproducer, these will not give optimum resolution because of spurious images due to line structure (Mertz and Gray, 1934; Wheeler and Loughren, 1938; Kell, Bedford and Fredenhall, 1940). The effect is particularly objectionable if the spot size is smaller than the line separation. For example, if the exploring spot at the transmitter is less than the width of the vertical line separation, the reproduction of a bright line nearly parallel to the scanning lines will be a row of separated bright beads, visible even at distances for which the line structure cannot be resolved. This and other spurious effects due to line structure can be minimized by making the distribution of the spot such that the response and brightness at the transmitting and receiving ends, respectively, are uniform over the entire field. A field, meeting this requirement, will be termed a " flat field ". Rectangular spots having a uniform distribution, whose height is equal to the line separation, will give a flat field.

In most cathode-ray systems the spot has radial symmetry, and a distribution such that the brightness or response is a maximum in the centre. To obtain a flat field in this case it is necessary to overlap the lines. The actual distribution across a scanning line may be represented by a cosine-squared distribution with a fair degree of accuracy. The minimum size of a spot having this distribution which will give a flat field requires 50 % overlap.

Band width. Resolution in the horizontal direction (along the scanning lines) is determined primarily by the frequency band width available in the communication channel. Signal frequencies generated in scanning a given image are determined by the rate at which the scanning spot moves across the image, and this, in turn, is dependent upon the number of scanning lines and the

picture-repetition rate. Thus it is at once apparent that an increase in the number of scanning lines in order to improve vertical resolution is accompanied by a corresponding decrease in horizontal resolution if the band width is fixed. Likewise an increase in picture-repetition rate, in order to reduce flicker, results in reduction in horizontal resolution. It is for this reason that interlaced scanning is used, since it gives a two-to-one increase in effective flicker frequency without the corresponding reduction in horizontal resolution.

Synchronizing signal. As was pointed out previously, the geometric fidelity of the picture reproduction depends upon the exact correspondence in position of the scanning spot at the transmitter and receiver. This correspondence requires that the periodicity and phasing of the horizontal and vertical scanning motions at the two termini be alike. It has not been found practicable to attempt

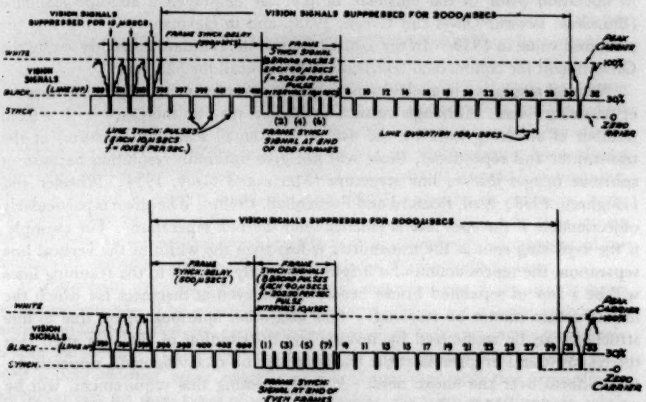

VISION SIGNALS SUPPRESSED FOR 10 μSECS.
VISION SIGNALS SUPPRESSED FOR 2000 μSECS
FRAME SYNCH DELAY (2000 μSEC)
FRAME SYNCH SIGNAL (5 BROAD PULSES EACH 90 μSECS f = 2424.25 PER SEC. PULSE INTERVALS 404 μSEC)
PEAK CARRIER
WHITE
VISION SIGNALS
BLACK
(4 LINES OFF)
SYNCH
100%
30%
0
ZERO CARRIER
LINE SYNCH PULSES (5 EACH 10 μSECS f = 10185 PER SEC.)
FRAME SYNCH SIGNAL AT END OF ODD FRAMES
LINE DURATION 101 μSECS
(1) (3) (5) (7)
(2) (4) (6)
FRAME SYNCH SIGNAL AT END OF EVEN FRAMES

Figure 13.

to maintain this condition with independent control of the scanning pattern at the pickup and viewing device. Instead, suitable signals are transmitted which indicate the beginning of each frame (or field repetition in the case of interlaced scanning) and the beginning of each line (Bedford and Smith, 1940). These synchronizing signals are part of the complete video signal, and occur during time not utilized by the picture signal itself, that is, during the interval in which the scanning spot is returning to its original position, after completing a line or field traversal. In terms of the frequency spectrum of the picture signal (discussed above) the frequency components of the synchronizing signal are multiples of line and frame frequency, their phase and amplitude being such that they have no effect on the picture except around the edges, where

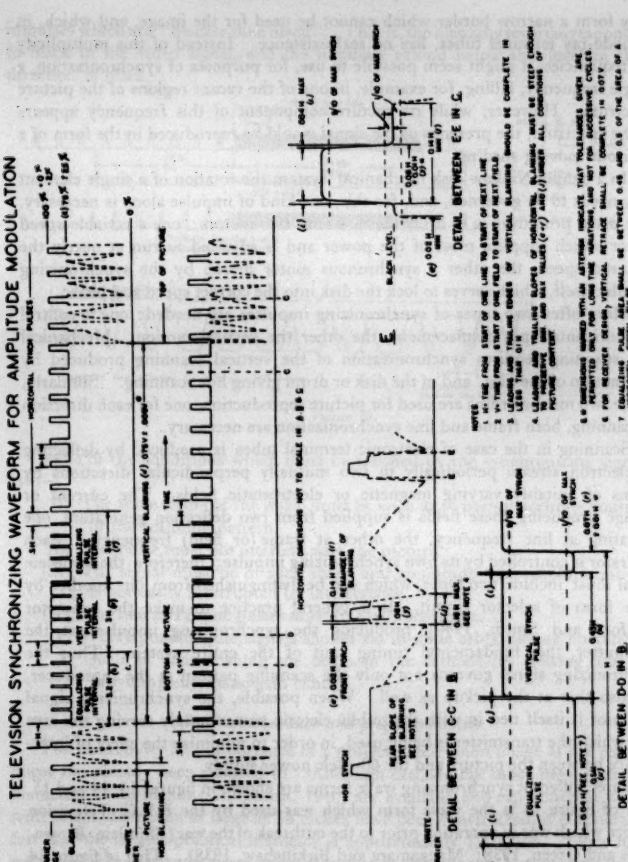

TELEVISION SYNCHRONIZING WAVEFORM FOR AMPLITUDE MODULATION
DETAIL BETWEEN C-C IN B.
DETAIL BETWEEN D-D IN B.
DETAIL BETWEEN E-E IN C.
NOTES
1 TIME FROM START OF ONE LINE TO START OF NEXT LINE
2 TIME FROM START OF ONE FIELD TO START OF NEXT FIELD
3 LEADING AND TRAILING EDGES OF VERTICAL BLANKING SHOULD BE COMPLETE IN LESS THAN 0.1H
4 LEADING AND TRAILING SLOPES OF HORIZONTAL BLANKING MUST BE STEEP ENOUGH TO MAINTAIN SYNC AND USE VALUES (0.4H) ARE (0.1H) WHEN ALL CONDITIONS OF PICTURE CONTENT
5 DIMENSIONS MARKED WITH AN ASTERISK INDICATE THAT TOLERANCES GIVEN ARE PERMITTED ONLY FOR LONG TIME VARIATIONS, AND NOT FOR SUCCESSIVE CYCLES.
6 FOR RECEIVER DESIGN, VERTICAL RETRACE SHALL BE COMPLETE IN 0.07V
7 EQUALIZING PULSE AREA SHALL BE BETWEEN 0.45 AND 0.5 OF THE AREA OF A HORIZONTAL SYNCH PULSE

Figure 14.

they form a narrow border which cannot be used for the image, and which, in cathode-ray terminal tubes, has no real existence. Instead of this multiplicity of frequencies, it might seem possible to use, for purposes of synchronization, a single frequency, falling, for example, in one of the vacant regions of the picture spectrum. However, while no picture component of this frequency appears in the spectrum, the presence of the signal would be reproduced in the form of a spurious moving shading.

In a simple Nipkow disk mechanical system the rotation of a single element only needs to be governed, and, for this, one kind of impulse alone is necessary. The usual procedure is to drive the disk with two motors: one a variable-speed motor which supplies most of the power and is adjusted to run at nearly the required speed, the other a synchronous motor driven by the synchronizing impulse itself, which serves to lock the disk into the correct speed and phase.

More often two types of synchronizing impulses are needed: one to control the horizontal spot displacement, the other the vertical motion. Mechanical film scanning requires synchronization of the vertical scanning produced by the motion of the film, and of the disk or drum giving line scanning. Similarly, where two mirror drums are used for picture reproduction, one for each direction of scanning, both frame and line synchronization are necessary.

Scanning in the case of electronic terminal tubes is produced by deflecting an electron stream periodically in two mutually perpendicular directions by means of suitably varying magnetic or electrostatic fields. The current or voltage producing these fields is supplied from two deflection generators, one operating at line frequency, the other at frame (or field) frequency. Each generator is controlled by its own synchronizing impulse; therefore the complete signal must include two types, which can be distinguished from one another by some form of selector circuit. It is general practice to make the generator (Bedford and Smith, 1940), producing the synchronizing impulses at the transmitter, the fundamental timing unit of the entire system. Thus the synchronizing signal governs not only the scanning pattern at the reproducer, but also that at the pickup as well. When possible, the synchronizing signal generator is itself tied in with the public electric power supply serving the area over which the transmitter is being used, in order to minimize the effect of interference between the picture and the 60-cycle power supply.

Three different synchronizing wave forms are shown in figures 13, 14 and 15. That of figure 13 is the wave form which was used by the B.B.C. Television Service which was in operation prior to the outbreak of the war (Blumlein, Brown, Davis and Green, 1938; Macnamara and Birkinshaw, 1938). That of figure 14 is the wave form recommended in the United States by the National Television System Committee, and that of figure 15 is an alternative wave form which is undergoing comparison tests with that of figure 14 in the United States.

Experience has prompted the almost universal adoption of synchronizing

impulses which are " blacker than black ". That is, the blanking level corresponds to black in the video signal, and the impulses extend below this level in the direction of black.

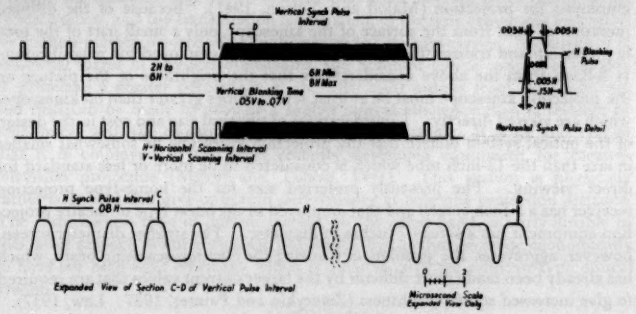


Figure 15.

Some of the conditions which must be satisfied by the synchronizing signal are:—

(i) The wave shape of the pulse must be such as to permit accurate control of the scanning pattern.
(ii) The pulse must not interfere with the picture.
(iii) The signal must be of such shape and amplitude that synchronization is not easily interrupted by interference.
(iv) The line and frame pulses must be easily separable.
(v) Where possible, the signals should aid in some other picture operation, such as extinguishing the beam in the cathode-ray viewing tube or providing automatic gain control.

(d) *Projection tube and projection optics*

The desirability and importance of a projected television image at the reproducer has already been referred to. Although cathode-ray tubes having screen diameters of 20 inches (Goldsmith, 1940) are available, and experimental tubes with screen diameters as large as 36 inches have been constructed, it is generally felt that the largest practical diameter is approximately 12 inches. Since it is very desirable to have pictures larger than this available for de-luxe type receivers in the home, and essential for theatre television, the use of optical magnification is the obvious expedient. Considerable work has been done on this problem, and satisfactory lenses have been found for both home-television receivers and theatre television.

In order to obtain adequate brightness in the projected image, it is necessary

that the brightness on the screen of the projection tube employed be great enough not only to offset the loss in brightness per unit area due to magnification, but also to compensate for losses due to the inefficiency of the optical arrangement employed for projection (Maloff and Tolson, 1941). Because of the diffused radiation of light from the surface of the kinescope, only a small part of the total is picked up and transmitted by the optical system employed for magnification. It follows from the above considerations that the brightness of the picture on the projection kinescope must be at least several times greater than on kinescopes which are viewed directly. Considerations of physical size and cost in the design of the optical system dictate that the projection kinescope be somewhat smaller in size than the 12-inch tube which is considered to be more or less standard for direct viewing. The presently preferred size for the home-type projection receiver has a 5-inch screen and that employed in the latest type of theatre projection equipment has a screen 7 inches in diameter. The smaller diameter screen, however, aggravates the problem of focusing the electron-scanning beam, which has already been made more difficult by the larger current values that are required to give increased screen brightness (Zworykin and Painter, 1937; Law, 1937).

An important problem in the development of projection-type kinescopes has been the performance of the fluorescent material, under the increased current and higher velocity conditions called for in this tube. Depending upon the secondary emissive properties of the screen material and other factors tied up with its chemical content, a point is reached where voltage saturation of the screen sets in and an increase in voltage beyond this point yields no appreciable increase in light output. Another important factor is the deterioration of the fluorescent material under bombardment by the high-intensity cathode-ray beam. Considerable progress has been made within the past few years on the problem of suitable screens for use in projection kinescopes, and it is to be expected that work now in progress will yield still further improvements (Zworykin and Painter, 1937; Zworykin and Morton, 1940).

It will be readily understood from the foregoing considerations that the optimum physical size of the projection kinescope represents a compromise between spot resolution, total light output from the fluorescent screen, and the physical size and cost of the optical system used for projecting the image. As further advances are made in electron-gun design (Law, 1937) and in improved efficiency of fluorescent materials it is to be expected that the diameter of the projection kinescope may be decreased, with a corresponding decrease in cost of the optical system, or that the projected images may be made still brighter.

At the present time projection receivers for home use employ potentials up to approximately 30,000 volts on the projection kinescope while, in television systems for theatre projection, voltages up to approximately 70,000 volts are utilized. At 70,000 volts the projection kinescope employed in the theatre-projection system is capable of delivering about 1200 lumens of useful light. With an overall optical efficiency of about 25 %, and with a screen having a

directional gain of two to one, a high-light brightness of slightly more than two foot-lamberts was obtained on a screen 15 × 20 feet (Maloff and Tolson, 1941).

Two general types of optical system are in use for projection of kinescope images: the reflection type of lens system, sometimes referred to as the Schmidt optical system, and the transmission lens system, of the type commonly employed for motion-picture projection. Each system has its advantages. Although it is somewhat larger in physical dimensions, the Schmidt optical system is capable of gathering a greater percentage of the total light output from the kinescope and thus can be made to provide a higher optical efficiency than is feasible or economical with the transmission type of optical system. It is for this reason that it is employed in the most recently developed equipment for theatre-television projection (Maloff and Tolson, 1941). This system is physically rather large and will give optimum performance at only one projection distance. The transmission type of optical system, on the other hand, is fairly flexible in this latter respect, is not very large physically, but cannot be made economically to have the high-light efficiency which is possible with the reflector-type optical system.

§ 8. CONCLUSIONS

(a) *F.C.C. findings*

On May 3, 1941, the U.S. Federal Communications Commission issued Orders, Rules and Regulations covering television broadcasting in the United States. These provided for the issue of commercial licences for television-broadcast stations, effective July 1, 1941. Standards covering the type of signal to be permitted in such operation were set up, so that the design, manufacture and sale of television receivers to the public might go ahead, with assurance that such receivers would not become obsolete in a short time. The standards were based upon recommendations made to the F.C.C. by the National Television Systems Committee, a special group set up in 1940 under the joint sponsorship of the Radio Manufacturers Association and the F.C.C. for the purpose of studying all known television systems and recommending standards believed to be capable of permitting the best possible operation (D. G. F., 1941).

The standards call for a 6 Mc. channel in which both the sound and picture signal must be transmitted, the spacing between the two unmodulated carriers being 4·5 Mc. and the sound carrier being located 0·25 Mc. below the upper extremity of the channel. Frequency modulation with a deviation of plus or minus 75 kc. is specified for the audio signal, and vestigial side-band operation for the video signal. The scanning standards call for 525 lines, with 2 to 1 interlacing, a field frequency of 60 per second and a frame frequency of 30 per second. Negative polarity is standard for the radiated picture signal (i.e., a decrease in initial light intensity causes an increase in radiated power). It is specified that the synchronizing signals and picture signals are to be radiated on the same carrier, and the synchronizing wave form is restricted to types which

will adequately operate a receiver responsive to the synchronizing wave form shown in figure 19. Horizontal polarization is made standard for both audio and video radiations (Brown, G. H., 1940). Other standards are given which are either necessary or desirable for a public television service (D. G. F., 1941).

In setting up the television broadcasting service, the F.C.C. made available 18 channels, each of 6 Mc. width. These channels do not comprise a continuous band, but are interspersed with other services in the radio spectrum between 50 Mc. and 294 Mc. To date, assignments for broadcasting have been made only in the lower eight channels, since presently available transmitting and receiving equipment performs most efficiently at the lower frequencies. Some of the upper channels are available for television-relay purposes under experimental licences in conjunction with the operation of commercial television-broadcast stations.

(b) *Present status of television broadcasting*

As this *Report* is written (August 1941), the United States is the only country, so far as we know, in which, as a result of the war, there is any activity in television broadcasting as a public service. In this country one station (the N.B.C. statio in New York City) is broadcasting television programmes on a commercial basis; several other stations are operating on a regular schedule, and some of these have announced their intention of beginning commercial operation within a few weeks. At the time of issuing the F.C.C. " Rules and Regulations governing Television Broadcasting Stations ", in May 1941, the F.C.C. also issued two lists of potential commercial stations: one list of five existing experimental television stations which would be eligible for commercial television-broadcast operation upon the filing and approval of an application for a construction permit, and a second list of sixteen experimental television stations which would be eligible for commercial operation upon the filing and approval of an application for a modification of construction permit. Thus it is seen that a fairly large number of potential commercial television stations were available in the United States on July 1, 1941.

The manufacturers of television receivers are now modifying sets which were sold to the public prior to the announcement of the present commercial standards, so that these receivers will operate in accordance with the new standards. A reawakening of public interest in television is beginning to manifest itself, and it is expected that the demand for television receivers will go up sharply as the television programme service is expanded. There is, however, considerable doubt, on account of the existing national emergency, as to whether the manufacturers of receivers will be able to provide receivers in large quantities. The war emergency has produced both a shortage in certain materials required in the manufacture of receivers and a shortage in the type of skilled personnel required to build the receivers. This situation will also seriously affect the manufacture of new equipment for television broadcasting stations. Although it seems certain

that these circumstances will have a retarding effect upon the expansion of television braodcasting in the United States, there is no indication at present that it will be necessary to suspend this service completely, as was done in England at the outbreak of the war.

(c) *Future (immediate) trends of television broadcasting*

As was previously indicated, the immediate future of television broadcasting is made somewhat uncertain by present war conditions, but it is the belief of the overwhelming majority of those persons who have had an opportunity to become familiar with the present and future possibilities in this new form of mass communication that it is destined to become the most important medium of home entertainment and education. There is no longer any question about the ability of television to render a service which will be welcome in the home, and for which people will be willing to buy receivers. Under normal conditions, it would be possible to predict with some certainty the rate at which this home service would be built up, but present world conditions make it difficult to determine exactly when such a service will be free to develop on a large scale. However, a television service comparable in extent to the present-day sound broadcasting service seems inevitable in the not too distant future.

It is possible that the intensive research and development work in radio for military purposes now going on all over the world will help provide new tools which will permit of an improved television service in future years. High power at higher carrier frequencies, with broader channels permitting greater picture definition and pictures in natural colour with adequate definition, are the benefits which it is reasonable to expect from this work. Networks of radio-relay stations connecting television-braodcast stations spread over large areas should be available within a relatively short time when the men and material can be made available for their planning and construction.

The projection-type receiver already developed is a clue to what the home television receivers of the near future will be like. This receiver will provide a bright picture of high definition and of sufficient size to be viewed at any comfortable viewing distance in the home. The same receiver cabinet in the future may provide for a facsimile reproducer and, as at present, will probably provide reception for standard and short-wave sound broadcasting bands, including the high-fidelity F.M. band. In the not too far distant future it is probable that the same television receiver will provide for reception in natural colour as well as in black and white, but our experience to date indicates that satisfactory colour will come only in channels substantially wider than the present standard 6 Mc. channels. It is still too early to predict with certainty whether the colour system of the future will be a simultaneous three-colour process or a cyclic system.

The successful demonstration of theatre television indicates a bright future

for this branch of the art. It does not seem too rash to predict that within a matter of a few years most local theatres, at least in the large metropolitan areas, will be equipped with television projection equipment and serviced by special television programmes over wire networks connecting to centralized distribution points. Theatres relying solely upon television may come into existence, but it seems more likely that television will be supplementary to the film projectors in existing theatres. The possibility of televising important events as they occur and distributing the television pictures to audiences in many theatres at remote points seems very attractive. Technically, the means for accomplishing this have been developed. Progress toward realization of such a service would seem to depend entirely upon the availability of equipment for doing the job, and this also is tied up with the present war conditions.

ACKNOWLEDGMENTS

Figures 3, 4, 5, 6, 10, 11, and 12 and portions of the text in the sections on *amplification and distribution*, *home receivers*, *the iconoscope*, and *scanning and synchronization*, are reprinted, by permission, from the book *Television*, by Zworykin and Morton, published by John Wiley and Sons, Inc., New York, N.Y.

The authors also wish to acknowledge the helpful assistance of Mr. F. J. Somers in connection with the preparation of the bibliography and of the section on *colour television*.

REFERENCES

ARMSTRONG, EDWIN H., 1936. *Proc. I.R.E.* **24**, 689, "A Method of reducing Disturbances in Radio Signalling by a System of Frequency Modulation".

BALDWIN, M. W., 1940. *Proc. I.R.E.* **28**, 458, "The Subjective Sharpness of Simulated Television Images".

BARTLETT, G. and CALDWELL, P. S., 1941. *General Electric Review*, **44**, 185, "Color Television".

BEDFORD, A. V. and FREDENHALL, G. L., 1939. *Proc. I.R.E.*, "Transient Response of Multistage Video Frequency Amplifiers".

BEDFORD, A. V. and SMITH, J. P., 1940. *R.C.A. Review*, **5** (no. 1), 51, "A Precision Television Synchronizing Signal Generator".

BEERS, G. L., SCHADE, O. H. and SHELBY, R. E., 1940. *Proc. I.R.E.* **28**, 430, "The R.C.A. Portable Television Pickup Equipment".

BLUMLEIN, A. D., BROWN, C. O., DAVIS, N. E. and GREEN, E., 1938. *J. Instn. Elect. Engrs.* **83**, 758, "The Marconi-E.M.I. Television System," Parts I, II and III.

BROWN, G. H., 1940. *Electronics*, **13** (no. 10), 20, "Vertical versus Horizontal Polarization".

BROWN, G. H., 1941. *R.C.A. Review*, **5** (no. 3), 301, "A Vestigial Sideband Filter for use with a Television Transmitter".

BROWN, W. J., 1939. *Electronics*, **12** (no. 12), 26, "British Television Receivers, I".

BROWN, W. J., 1940. *Electronics*, **13** (no. 3), 26, "British Television Receivers, II".

DEVORE, H. B. and IAMS, H., 1940. *Proc. I.R.E.* **28**, 369, "Some Factors affecting the Choice of Lenses for Television Cameras".

D. G. F., 1939–41. See F., D. G., 1939–41.

EDDY, WILLIAM C., 1940. *R.C.A. Review*, **4** (no. 4), 414, "Television Lighting".

ENGSTROM, E. W., 1939. *J. Appl. Phys.* **10**, 455, "Television Receiving and Reproducing Systems".

ENGSTROM, E. W. and HOLMES, R. S., 1938. *Electronics*, **11** (no. 4), 28; **11** (no. 6), 20; **11** (no. 8), 18; **11** (no. 11), 18, "Television Receivers".

ENGSTROM, E. W. and HOLMES, R. S., 1939. *Electronics*, **12** (no. 1), 19; **12** (no. 4), 22, "Television Receivers".

ESPENSCHIED, L. and STRIEBY, M. E., 1934. *Elect. Engng.* **53**, 1371, "Wide Band Transmission over Coaxial Lines".

FARNSWORTH, P. T., 1934. *J. Franklin Inst.* **218** (no. 4), 411, "Television by Electron Image Scanning".

F., D. G., 1939. *Electronics*, **12** (no. 3), 26, "Television Transmitters".

F., D. G., 1940. *Electronics*, **13** (no. 10), 32, "Color Television demonstrated by C.B.S. Engineers".

F., D. G., 1941. *Electronics*, **14** (no. 4), 18, "Groundwork laid for Commercial Television".

FINK, D. G., 1940. *Principles of Television Engineering* (New York: McGraw-Hill).

FINK, D. G., 1941. *Proc. I.R.E.* **29**, 310, "Brightness Distortion in Television".

FOSTER, D. E. and RANKIN, J. A., 1941. *R.C.A. Review*, **5** (no. 4), 409, "Video Output Systems".

GODDARD, D. R., 1939. *Proc. I.R.E.* **27**, 692, "Transatlantic Reception of London Television Signals".

GOLDMAN, S., 1939. *Proc. I.R.E.* 27, 725, "Television Detail and Selective Sideband Transmission".

GOLDMARK, P. C., 1939. *J. Appl. Phys.* **10**, 447, "Problems of Television Transmission".

GOLDMARK, P. C., 1940. *Electronics*, **13** (no. 12), 26, "Color Television" (I.R.E. Rochester Convention).

GOLDMARK, P. C. and DYER, J. N., 1940. *Proc. I.R.E.* **28**, 343, "Quality in Television Pictures".

GOLDSMITH, T. T., 1940. *Electronics*, **13** (no. 6), 16, "A 20-inch Television Receiver".

HANSON, O. B., 1937. *R.C.A. Review*, **1** (no. 4), 3, "Experimental Studio Facilities for Television".

HANSON, O. B., 1941. *R.C.A. Review*, **5** (no. 3), 267, "R.C.A.-N.B.C. Television presents a Political Convention as First Long Distance Pickup".

HEROLD, E. W., 1938. *Communications Magazine*, **18** (no. 8), 11, "High Frequency Correction in Resistance Coupled Amplifiers".

HEROLD, E. W., 1940. *R.C.A. Review*, **4** (no. 3), 324, "Superheterodyne Converter Considerations in Television Receivers".

HOLLYWOOD, J. M., 1939. *Proc. I.R.E.* **27**, 457, "Single Sideband Filter Theory with Television Applications".

IAMS, H., MORTON, G. A. and ZWORYKIN, V. K., 1939. *Proc. I.R.E.* **27**, 541, "The Image Iconoscope".

IVES, H. E., 1931. *J. Opt. Soc. Amer.* **21**, 2, "Television in Colors from Motion Picture Film".

IVES, H. E. and JOHNSRUD, A. L., 1930. *J. Opt. Soc. Amer.* **20**, 11, "Television in Colors by a Beam Scanning Method".

JANES, R. B. and HICKOK, W. H., 1939. *Proc. I.R.E.* **27**, 535, "Recent Improvements in the Design and Characteristics of the Iconoscope".

JENSEN, A. G., 1941. *Proc. I.R.E.* **29**, 243, "Film Scanner for Use in Television Transmission Tests".

KALLMAN, H. E. and SPENCER, R. E., 1940. *Proc. I.R.E.* **28**, 257, "Transient Response of Single Sideband Systems".

KELL, R. D., BEDFORD, A. V. and FREDENHALL, G. L., 1940. *R.C.A. Review*, **5** (no. 1), 8, "A Determination of Optimum Number of Lines in a Television System".

KELL, R. D. and FREDENHALL, G. L., 1940. *R.C.A. Review*, **4** (no. 4), 425, "Selective Sideband Transmission in Television".

KROGER, F. H., TREVOR, B. and SMITH, J. E., 1940. *R.C.A. Review*, **5** (no. 1), 31, "A 500-Mc. Radio Relay Distribution System for Television".

LARNER, E. T., 1929. *Practical Television* (New York: D. Van Nostrand Co.).

LAW, R. R., 1937. *Proc. I.R.E.* **25**, 954, "High Current Electron Gun for Projection Kinescopes".

LINDENBLAD, N. E., 1939. *R.C.A. Review*, **3** (no. 4), 387, "Television Transmitting Antenna for Empire State Building".

LOHR, L., 1940. *Television Broadcasting* (New York: McGraw-Hill).

LUCKEISH, M., 1927. *Color and its Application* (New York: D. Van Nostrand Co.).

MCILWAIN, K., 1939. *J. Appl. Phys.* **10**, 432, "Survey of Television Pickup Devices".

MACNAMARA, T. C. and BIRKINSHAW, D. C., 1938. *J. Instn. Elect. Engrs.* **83**, 729, "The London Television Service".

MALOFF, I. G., 1939 *R.C.A. Review*, **3** (no. 4), 409, "Gamma and Range in Television".

MALOFF, I. G. and TOLSON, W. A., 1941. *R.C.A. Review*, **6** (no. 1), 5, "A Résumé of the Technical Aspects of R.C.A. Theatre Television".

MERTZ, P., 1939. *J. Appl. Phys.* **10**, 443, "High Definition Television".

MERTZ, P. and GRAY, F., 1934. *Bell Syst. Tech. J.* **13**, 464, "Theory of Scanning".

MORRIS, R. M. and GUY, R. F., 1940. *R.C.A. Review*, **5** (no. 2), 190, "N.B.C. Frequency Modulation Field Tests".

MORRIS, R. M. and SHELBY, R. E., 1937. *R.C.A. Review*, **2** (no. 1), 14, "Television Studio Design".

MOUNTJOY, G., 1940. *R.C.A. Review*, **4** (no. 3), 299, "Simplified Television I.F. Systems".

NERGARD, L. S., 1939. *Proc. I.R.E.* **27**, 666, "Theoretical Analysis of Single Sideband Operation of Television Transmitters".

POCH, W. J. and EPSTEIN, D. W., 1937. *Proc. I.R.E.* **25**, 15, and *R.C.A. Review*. **1** (no. 3), 19, "Partial Suppression of One Sideband in Television Reception".

PROTZMAN, A. W., 1940. *R.C.A. Review*, **4** (no. 4), 399, "Television Studio Technique".

ROBINSON, D. M., 1939. *Proc. I.R.E.* **27**, 483, "The Supersonic Light Control and its Application to Television, with Special Reference to the Scophony Television Receiver".

ROSE, A. and IAMS, H., 1939 a. *Proc. I.R.E.* **27**, 547, "Television Pickup Tubes using Low Velocity Beams".

ROSE, A. and IAMS, H., 1939 b. *R.C.A. Review*, **4** (no. 2), 186, "The Orthicon, A Television Pickup Tube".

ROSENTHAL, A. H., 1940 a. *Proc. I.R.E.* **28**, 203, "A System of Large Screen Television Projection based on certain Electron Phenomena in Crystals".

ROSENTHAL, A. H., 1940 b. *Electronics and Television and Short Wave World*, **13**, 52 and 117, "The Skiatron".

SARNOFF, D., 1939. *J. Appl. Phys.* **10**, 426, "Probable Influences of Television on Society".

SEELEY, S. W., 1941. *R.C.A. Review*, **5** (no. 4), 468, "Frequency Modulation".

SMITH, T. A., 1940. *R.C.A. Review*, **4** (no. 3), 290, "R.C.A. Television Field Pickup Equipment".

STRIEBY, M. E. and WEIS, C. L., 1941. *Proc. I.R.E.* **29**, 371, "Television Transmission".

TILTMAN, R. F., 1928. *Radio News*, **10** (no. 4), 320, "Television in Natural Colors demonstrated".

TYSON, B. F., 1940. *Electronics*, **13**, (no. 11), 23, "A Preselector Circuit for Television".

WENTZ, J. F., 1938. *Bell Laboratories Record*, **16**, 196, "Transmission Characteristics of the Coaxial Structure".

WHEELER, H. A., 1939 a. *Proc. I.R.E.* **27**, 359, "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes".

WHEELER, H. A., 1939 b. *Proc. I.R.E.* **27**, 429, "Wide Band Amplifiers for Television".

WHEELER, H. A. and LOUGHREN, A. V., 1938. *Proc. I.R.E.* **26**, 540, "The Fine Structure of Television Images".

ZWORYKIN, V. K., 1928. U.S. Patent No. 1,691,324, issued November, 1928.

ZWORYKIN, V. K., 1929. *Radio Engng.* **18** (no. 8), 38 "Television with Cathode Ray Receiver".

ZWORYKIN, V. K., 1938. U.S. Patent No. 2,141,059, issued December, 1938.

ZWORYKIN, V. K. and MORTON, G. A., 1940. *Television* (New York: John Wiley and Sons; London: Chapman and Hall).

ZWORYKIN, V. K., MORTON, G. A. and FLORY, L. E., 1937. *Proc. I.R.E.* **25**, 1071, "Theory and Performance of the Iconoscope".

ZWORYKIN, V. K. and PAINTER, W. H., 1937. *Proc. I.R.E.* **25**, 937, "Development of Projection Kinescope".

INSTRUMENTAL TECHNIQUE IN ASTROPHYSICS

By A. HUNTER,
Royal Observatory, Greenwich

§1. TELESCOPES AND SPECTROGRAPHS

PROMPTING many of the recent technical improvements in astrophysical apparatus has been the struggle for more and more light, with the double object of probing ever farther into the depths of space and (less spectacular but no less important) of studying more closely our nearer galactic neighbours. The most obvious way to collect more light is, of course, to increase the collecting area. Doubling the aperture of a telescope is easy enough on paper, but the optical and engineering problems involved are much more than doubled in the attempt. There is no need to describe here how these problems have been tackled, and successfully overcome, in building the 200-inch telescope for Mt. Palomar from the experience accumulated in the building and operation of the 100-inch telescope on Mt. Wilson: this story formed the subject of the twelfth Thomas Young Oration delivered by the Astronomer Royal (Jones, 1941). The completed dome now houses the mounting ready for use, but it is now announced that the final figuring of the mirror, always a tedious business because of the lengthy delays necessary between polishing and testing, has been interrupted by the exigencies of the U.S. defence programme. European astrophysicists have plenty of reason for sympathising with their American colleagues in this setback.

The 200-inch telescope will spend much of its time in feeding a nebular spectrograph, in order that the spectra of faint external galaxies may be studied. The main requirement in such spectrographs is speed—speed high enough, if need be, to overcome the reflection losses at many glass-air surfaces. This requirement is met by a remarkable short-focus camera objective designed by Bracey (1936) and working at f/0·36. The focal plane of this lens lies on the rear surface of its last component, with which the photographic plate is kept in contact by an immersion fluid. Other such high-speed camera lenses have also been designed, and now that the coating of glass surfaces with evaporated fluorite for the removal of reflection losses is an accomplished fact (Cartwright, 1940) we may expect first that the performance of existing spectrographs can be improved by a factor of at least two; and second, what is perhaps more important, that designers will be relieved in future of the nagging fear of too many internal reflecting surfaces. In any case, opticians have now provided the astrophysicist

with instruments which are already capable of recording in a few minutes the spectrum of a slit-full of light from the night-sky background. It must not, however, be supposed that, this natural limit having once been reached, no further progress is possible. Already the width of slit in the most modern nebular spectrographs is so adjusted that its monochromatic image on the plate is equal in breadth to the resolving power of the emulsion; but of course slower, finer-grain plates would be preferred if exposures were not thereby increased. Further, a greater dispersion would be of advantage for radial-velocity measurements, so that there is still a field in which designers may compete to produce an instrument capable of reaching the common natural limit in the shortest possible exposure time.

A recent competitor for this distinction is the nebular spectrograph (figure 1) designed by Baker (1940 a) and now being constructed for the Harvard College

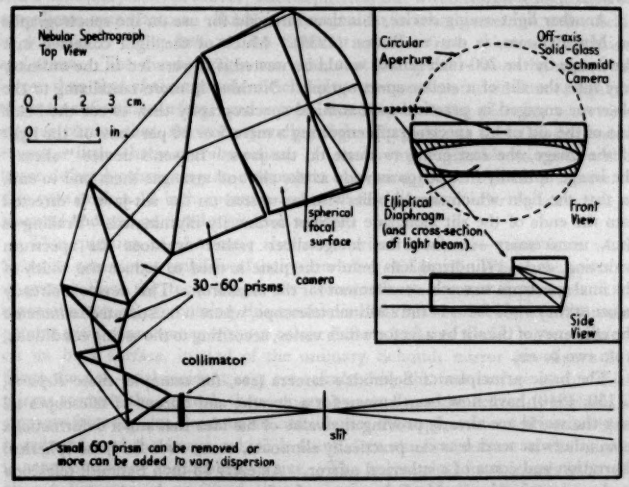


Figure 1. Nebular spectrograph employing a solid-glass Schmidt camera.

Observatory. The light enters the slit at f/5 and is rendered parallel by a 10 mm. f/5 cemented achromat, coated on both external surfaces. It is then dispersed by one or more 60° prisms, followed by two, three, or four 30°–60° prisms, all with coated surfaces, which add dispersion and spread the collimated circular beam to one of elliptical cross-section, which then enters the camera. This latter is of novel design: it is a solid-glass Schmidt camera working at the extreme

focal ratio of 0·30. It is used off the axis in a direction perpendicular to the dispersion, so as to obtain access to the spherical focal surface. Recording is done on ordinary 8 mm. roll-film fed between spools and pressed against the focal surface by means of a plunger. Oil immersion is necessary in order that the light may finally emerge from the glass on to the film. The working range, which is covered normally in two settings, is from 3900 A. to 6700 A., with a linear dispersion variable at will between 500 and 250 A./mm. at Hβ. By suitably stopping down the camera aperture to f/0·60, the entire range of the spectrum may be brought into reasonable focus simultaneously. Figure 1 shows that the spectrograph is remarkably compact, which makes for ease of mounting and of temperature control. The prisms are small and inexpensive, and lose little light by absorption; the camera, though difficult to construct, makes up for this by its extraordinary speed. Results given by this instrument will be awaited with interest.

Another light-saving device, this time intended for use on the spectrographs on Mt. Palomar, is due to Bowen (1938). Much of the light collected into the image by the 200-inch mirror would be wasted if it were fed in the ordinary way into the slit of a stellar spectrograph. Nothing is more tantalizing to the observer engaged in practical astronomical spectrography than to see the black line of the slit of his spectrograph engulfing a mere 5 or 10 per cent of the light of the image, the rest going to waste on the jaws. Bowen's device "slices" the image optically into strips as wide as the slit and arranges them end to end, so that the light which would otherwise be wasted on the slit-jaws is directed into the ends of the slit alongside the part ordinarily illuminated. Trailing is thus unnecessary—in fact the image-slicer rather overdoes the spectrum widening, and a cylindrical lens before the plate is used to reduce the width of the final spectrum to a value convenient for the measurer. This device is already in use at the coudé focus of the 100-inch telescope, where it is estimated to increase the efficiency of the slit by a factor which varies, according to the seeing conditions, from two to six.

The basic principles of Schmidt's camera (see, for example, these *Reports*, **7**, 130, 1940) have now been known for a decade, and dozens of telescopes all over the world are already proving the value of his idea that small deformations in an otherwise weak lens can practically eliminate, over a wide field, the spherical aberration and coma of a spherical mirror. An early 18-inch Schmidt telescope has been in operation on Mt. Palomar now for five years, employed in a systematic search for supernovae in external galaxies. Enough bright ones have been discovered to make profitable a detailed study of their spectra (Minkowski, 1939), with results that were puzzling enough at the time (see these *Reports*, **6**, 137, 1939), but now seem to be yielding to a line of attack which is unique in astrophysics. This consists in matching the observed spectra by synthetic ones formed by combining ordinary line spectra, each with suitable intensity and plausible Doppler widening. The early results of this process (Payne-Gaposchkin and Whipple, 1940) are distinctly promising.

An outstanding disadvantage of the Schmidt camera is that the star images come to a focus on a spherical surface of radius equal to the focal length. For short-focus cameras, photographic plates cannot be bent safely to the required curvature, and a large area of film is not altogether satisfactory. For the 48-inch Schmidt telescope for Mt. Palomar, Ross (1940) has computed a field-flattening plano-convex lens placed close to the focus. This somewhat increases the chromatic aberration (which for the Schmidt correcting plate alone is well within the photographic tolerance), but not to an objectionable degree. If necessary, in future work, the power of the field-flattening lens can be reduced so as to leave a residual curvature which can be taken up by bending the photographic plate. Another way of removing field curvature which may prove fruitful is described by Baker (1940 b). The camera he suggests consists of a correcting plate similar to the Schmidt, together with a primary concave mirror and a secondary convex mirror, both aspherical but without lines of inflexion to complicate manufacture. Such an arrangement is anastigmatic on a flat field, in addition to being free, as is a Schmidt camera, from spherical aberration and coma. The disadvantages are, first, that the aperture must be somewhat larger than that of the equivalent Schmidt to compensate for the loss of light at the extra reflecting surface; and second, that the correcting plate has somewhat steeper curves than the corresponding Schmidt plate, so that the dispersion causes a little chromatic aberration. It is, however, possible to remove this latter defect by designing an achromatic correcting plate. An f/2·4 camera of this type with a 12-inch aperture and a 7° field is now under construction at Harvard and will be fitted with such a plate.

This question of chromatic aberration arises in a more acute form when Schmidt cameras of small focal ratios are being designed. Where the aperture is greater than the focal length the colour effects become appreciable. An ingenious way of overcoming this trouble is described by Hendrix (1939), who obtains great convergence without colour defects by using a solid mirror silvered on its back surface, instead of the ordinary Schmidt mirror (see figure 2*a*). Since the non-axial incident rays are refracted by the first surface of this thick mirror towards the normal, they appear, as seen from the reflecting surface, to emanate from a point closer to the axis; hence the angle subtended by an object is reduced, and the image, which is focused on the first surface of the mirror, is correspondingly diminished in size. The geometrical focal length is hardly altered, however, so an f/0·66 camera can be constructed with field curvature and correction-plate curvatures no greater than those of an f/1·0 Schmidt of the ordinary type.

This principle may be carried still further: the correcting surface may be put at the focus (though this introduces manufacturing difficulties), or it may be arranged that there is no medium other than glass between the correcting surface and the focus, either (figure 2*b*) by putting the photographic plate in a hole in the thick glass, or (figure 2*c*) by using a " folded " solid type (Hendrix

and Christie, 1939) in which the incident beam is reflected through 90° by a 45° totally reflecting surface midway in the camera, the focal surface being made accessible by a small right-angle prism cemented to the 45° face. Several of these folded-type telescopes are in successful operation for astrophysical work. A solid-type camera in which easy access is arranged to the focal surface by using off-axis rays has been described above (see figure 1).

All such modifications to the original Schmidt telescope suffer from the defect that homogeneous non-absorbent glass is needed for the solid component. For work in the extreme ultraviolet they will therefore be at a disadvantage, apart altogether from the difficulty of manufacture, and it must not be supposed that the supremacy of the original design for spectrographic use is directly threatened. For instance, the coudé spectrograph of the 100-inch telescope has recently been completely redesigned (Adams, 1941) in order to incorporate

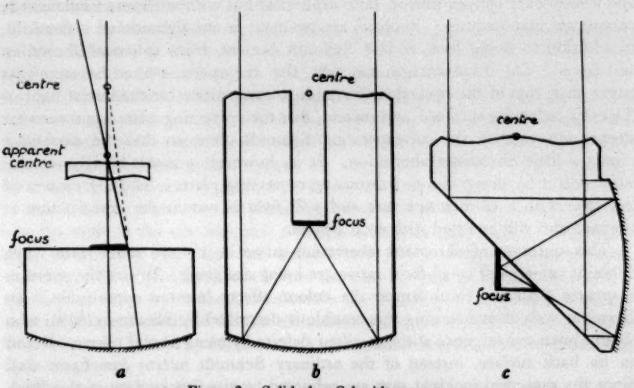


Figure 2. Solid-glass Schmidt cameras.

three ordinary Schmidt cameras with focal lengths of 32, 73, and 114 inches respectively, the dispersing device being a plane diffraction grating ruled in aluminium on Pyrex glass. Not the least of the recent advances in instrumental technique has been the remarkably successful development of such gratings by Prof. R. W. Wood. Ruling with diamond points of different shapes on aluminium surfaces, Wood is able to concentrate most of the incident light (in some cases 85 per cent) into almost any desired order. This technique has been developed to such an extent that the grating is now a serious rival of the prism in stellar work. In the case of this spectrograph the concentration is in the red and infra-red of the first order and the violet and blue of the second. Since the reflecting surfaces involved, both in the main telescope and in the spectrograph, are aluminized, high ultraviolet transmission is obtained; and an

important practical point is that ultraviolet iron lines of the second order may be used as comparison lines in regions of the spectrum where the first-order iron lines are few in number. The performance of the instrument is extraordinary. The spectrum of a star near the limit of naked-eye visibility is obtained with the 114-inch Schmidt camera (second order, 2·9 A./mm.) in 5½ hours. Such a dispersion, particularly in the red end of the spectrum, is impossible with stellar prism spectrographs, which are, moreover, useless in the ultraviolet. The time is evidently now at hand when the spectrum of a bright star may be studied in almost as much detail as that of the sun. Large linear dispersion is most valuable for the accurate determination of radial velocities and the physical study of line contours. Already the apparatus at Mt. Wilson has been used (Adams, 1941) for a provisional determination of the solar parallax from spectrograms of Arcturus; for a study of the spectrum of Mars, which as a result is now known not to show water-vapour lines unless their intensities are less than five per cent of those due to the earth's atmosphere; and for an investigation of some faint ultraviolet interstellar lines which has established the hitherto unsuspected existence in interstellar space of CH and CN molecules.

The revolution which was predicted in stellar spectroscopy by the perfection of the process for aluminizing glass surfaces has certainly come about. The main mirror of the Mt. Wilson 100-inch telescope has recently been re-aluminized (Humason, 1940) after the old surface had given five years' service at an ultraviolet efficiency quite unattainable with silvered surfaces. Experience at other observatories in America has been reviewed by Williams (1939), and it is certain that in all but the most industrialized areas (such as Greenwich, for example, where aluminized surfaces are found serviceable but by no means imperishable, and are therefore considered worth while only where resilvering is inconvenient or where ultraviolet reflectivity is vital) aluminizing has come to stay. It is sometimes overlooked, however, that the reflecting power of aluminium is lower in the yellow and red than that of silver, and that accordingly, where many reflections at aluminium surfaces are involved, a serious loss of light may occur in these regions. This means, for example, that the coudé spectrograph of the 100-inch telescope is not very efficient for the observation of faint or early-type stars in the red region of the spectrum (Adams, 1941). Nevertheless, such disadvantages are not insuperable, and the supremacy of aluminium is now so secure that the complicated plant necessary for aluminizing the 200-inch mirror is provided without question as an essential accessory in the Palomar equipment.

§ 2. PHOTOMETERS AND MICROPHOTOMETERS

Visual photometry in astrophysics is not unnaturally limited in application, though in following the rapid solar changes which now attract so much attention, simple visual apparatus (Newton, 1939) is proving useful. Photography has been, and still is, at the root of most astrophysical photometry, whether this is concerned with direct measurements of stellar magnitudes or with such spectrophotometric studies as temperature investigations. In colour work, however,

the complexity of the photographic process is against its use, and such effects as those due to variation of exposure time or to reciprocity failure in the emulsions employed make accurate work tedious or even impossible. Moreover, patchiness in the sensitivity of photographic emulsions appears unavoidable and sets a limit to the accuracy attainable even when, as is essential, each plate is separately standardized.

In such circumstances it is not surprising that where speed in obtaining results is a consideration, astrophysicists have been turning more and more to photoelectric photometry. Until quite recently the application of photoelectric methods at the telescope has been mostly to colour-index work; the light of a star, filtered through first one colour screen and then another, is admitted to the cathode of the photocell (usually one of the Cs–O–Ag type), and the cell's response noted in the two cases, often by measuring the rate of charge of the anode with a Lindemann electrometer. The great advantage of this method is the accuracy and reproducibility of the results obtained. But the process is not truly spectrophotometric, and the spectral regions measured are dictated more by the instrumental conditions than by the astrophysical requirements. For measurements at definite wavelengths, then, the photographic method, involving spectral isolation by auxiliary dispersing apparatus, has been preferred, accuracy being obtained by repetition at the expense of speed. It is not to be regarded as adverse criticism of the photoelectric results to remark that their physical interpretation is complicated (Atkinson, Hunter and Martin, 1940), chiefly by the fact that the spectral regions involved are wide and often overlap. Quite recently, however, the outlook has been changed by the development of a photoelectric method which is nearly true spectrophotometry (Hall, 1941). In this technique emphasis is placed on the spectral definition of the energy admitted to the cell, without sacrificing the speed and accuracy which are the hallmarks of the photoelectric method. Spectra are formed in the focal plane of the telescope by means of a wire or replica grating, and two movable slits isolate the wavelength regions to be admitted to the cell. The arrangement is thus in effect a monochromator backed by a photocell. Though it has been used as yet only in conjunction with a small-aperture telescope, the results suggest that in the spectra of bright stars, regions only 1 A. wide could be measured with a large reflector. Direct photoelectric study of stellar line contours may thus be possible in the fairly near future; and as it is, the speed of the apparatus already makes feasible a study of stellar colour as a function of the spectral region in which it is determined, and of such factors as invisible companions, interstellar absorption, and luminosity.

In spite of the rise of photoelectric methods, photographic photometry is still, of course, very widely used. Especially is this so in the study of line profiles, where theoretical progress has perhaps outstripped observational technique in the last decade. The procedure used in obtaining properly calibrated spectrograms is now fairly well standardized, and their measurement is in theory easy

enough with the aid of recording microphotometers. Nevertheless the process of reducing microphotometer tracings to relative intensity curves is in practice discouragingly long and monotonous, especially where widely different wavelengths are recorded on the same plate. The ordinary method of reduction applied to such a task, for instance, as that of reproducing Fraunhofer-line profiles from solar spectrograms has always been considered prohibitively tedious if the study is to be at all exhaustive. This is so even though the richness of the field of investigation which would thereby be opened up has been fully appreciated by astrophysicists. Recently, however, the position has been eased by the development of a method of automatic registration of true intensities, direct from the photometric plate (Minnaert and Houtgast, 1938). In this method a photographic characteristic curve connecting transmission T with intensity I is first drawn, for a given wavelength region, from the photometric scale markings on the plate. This is then cut out and used as a diaphragm in an ingenious attachment which will convert any microphotometer to one giving relative intensity traces. This attachment consists of an auxiliary galvanometer, whose mirror reflects a strip of light on to the diaphragm, and a lens and photocell behind the diaphragm to collect the fraction of incident light which it allows to pass. The photo-current generated by the analysing beam of the microphotometer (after its transmission through the spectrogram under examination) is diverted from the ordinary recording galvanometer to the auxiliary galvanometer, whose mirror is deflected by an amount proportional to T. The (T, I) diaphram, owing to its shape, allows only a fraction of the light proportional to I to reach the collecting lens and photocell behind it. The output of this photocell is supplied to the recording galvanometer of the microphotometer, which then records a direct intensity trace.

This apparatus has its defects—for example, mean (T, I) curves must be used over more or less wide wavelength intervals—but it has made possible for the first time the production of a precise photometric atlas of the solar spectrum (Minnaert, Mulders and Houtgast, 1940). This monumental work, the modern analogue of Rowland's photographic atlas, was completed, and a few copies were distributed, after the invasion of the Netherlands. It consists of intensity tracings from 3332 A. to 8771 A. reproduced from Mt. Wilson and Utrecht spectrograms on a uniform scale of 2 cm. per A. There are thus over 100 yards of tracings in the whole volume. A section in the region of the K line of Ca II is reproduced in figure 3 together with a photograph of the same region. Wavelengths can be read from the original tracings correct to 0·01 A., and it is estimated that central intensities and equivalent widths can be obtained to within five per cent. That intensities throughout the whole solar spectrum are now available to this accuracy is a great achievement, and many astrophysical problems hitherto considered insoluble will not seem so intractable with this atlas at hand.

The direct registration of intensities is made yet easier and more accurate by a still more recent development (Williams and Hiltner, 1940) in which no

stencil-cutting is required, nor are mean characteristic curves used at wavelengths to which they are not strictly applicable. The calibration spectrum impressed on each plate is here arranged to vary exponentially in intensity across the dispersion, and is scanned by one beam of light at the same wavelength as is the stellar spectrum, simultaneously, by another similar beam. The first beam, by moving across the direction of dispersion, is caused by suitable optical and electrical means to pick out at each wavelength the position on the calibration spectrum where the density is equal to that of the stellar spectrum at the same wavelength. This cross-motion is transferred, by means of a cam and mirror, to the motion of a spot of light across the recording paper, the cam being so shaped that the recording spot moves an amount which is proportional to the intensity of the light which was originally impressed on the calibration spectrum.

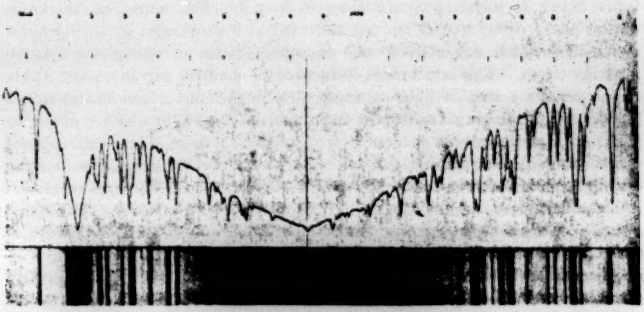


Figure 3. Section, in the region of the K line, of the photometric atlas of the solar spectrum.

Since calibration spectrum and stellar spectrum are kept together in both wavelength and density, the motion of the recording spot will be proportional to the intensity in the stellar spectrum; thus direct intensity tracings will result. The convenience of such an instrument as this will make immediately practicable large-scale spectrophotometric programmes which have until now been projects for a future hardly in sight.

§3. CORONAGRAPHS AND SPECTROHELIOGRAPHS

The one sphere in which the designer of astrophysical apparatus can work without encountering trouble from lack of sufficient light is solar research. In one particular problem the astrophysicist is indeed faced with too much light, viz., the direct sunlight, scattered by the earth's atmosphere and the optical parts of the telescope, which obscures the solar corona except during a total eclipse. The problem of photographing the solar corona or its spectrum outside eclipse, long considered impossible after many abortive attempts, has been

brilliantly solved in the last decade by the painstaking work of Lyot (1939). He reduces atmospheric diffusion of sunlight by observing at a high altitude (9500 ft.) ; diffusion in the optical parts by using the most flawless glass, with a polish as nearly perfect as possible; and diffraction at the edge of the object-glass by using a suitable masking diaphragm. In its most recent form Lyot's coronagraph is shown in figure 4. A flawless simple plano-convex lens 20 cm. in aperture and 4 m. in focal length (not shown, to the left) forms a solar image on a blackened brass disc A projecting 15 to 20 seconds of arc beyond the sun's limb. The field lens behind A produces an image of the objective on a diaphragm in the shadow beyond: the function of this diaphragm is to cut off the light diffracted at the edge of the first lens. After total reflection at D the light is focused by the achromat F via a further prism E either on to the slit G of a

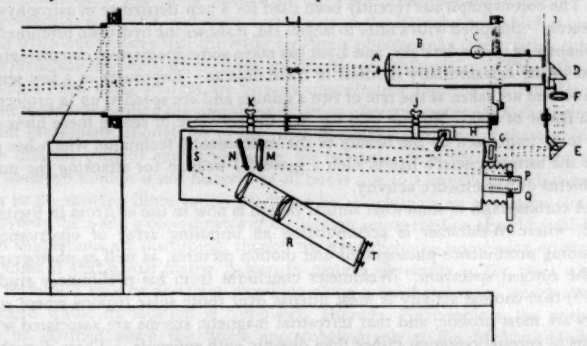


Figure 4. The coronagraph with spectrograph attached.

spectrograph or on to a photographic plate. For direct photography the lens A is turned on its axis during exposures so that particles of dust on its surface do not produce spots which can be mistaken for coronal features. In the spectrograph a plane grating can be placed at N to form an autocollimating arrangement, with the spectrum at O; or, alternatively, with the grating at S the camera R forms a shorter spectrum at T.

With the coronagraph alone, many photographs of the inner corona (up to 7′ from the limb) have been secured throughout the last solar maximum in 1937-38. The outer corona is too faint to be recorded, but the method, though inferior in this respect to that utilizing a total eclipse, has the enormous advantage that variations of the coronal streamers and arches with time can be followed. The existence of day-to-day variations, hitherto only suspected, has been established by this method. With the coronal spectrograph, the spectrum of the inner corona (40″ from the limb) has been explored from 3350 A. to 12000 A.

with a dispersion up to 5 A./mm. In the ultraviolet, violet, and blue, atmospheric diffusion makes the new method inferior to the old. In the green there is not much to choose between the results, but beyond about 5500 A. the new method is superior to eclipse observations, and three strong lines at, respectively, 7060 A. (since found on eclipse plates), 10747 A., and 10778 A. have actually been discovered by Lyot.* The plates are more numerous and on a larger scale than most eclipse spectrograms, so that accidental markings can be more certainly distinguished from true lines, and the wavelengths can also be measured with greater accuracy. Until much faster infra-red plates become available, the new method will be the only one even possible for studies of the coronal spectrum in the infra-red, and for following day-to-day variations in all spectral regions it is of course essential.

The coronagraph has recently been used for a new departure in astrophysical research. Supplied with a filter to isolate Hα, it shows the hydrogen prominences brilliantly against a dark sky, and Lyot has taken some spectacular cinematograph films of the extraordinary motions of these objects. Exposures of a few tenths of a second are taken at the rate of two a minute and are speeded up in projection by a factor of 600. No one who has had the privilege of seeing these films can fail to be impressed by the beauty of the instrumental technique which has put into the astrophysicist's hands such a powerful weapon for attacking the many problems of prominence activity.

A coronagraph of somewhat similar design is now in use at Arosa in Switzerland, where Waldmeier is accumulating an imposing array of observations, including prominence photographs and motion pictures, as well as photographs of the coronal spectrum. Waldmeier concludes from his preliminary studies (1939) that coronal activity is most intense over those solar regions where sunspots are most prolific, and that terrestrial magnetic storms are associated with bursts of coronal radiation rather than directly with sunspots. These disturbed regions of the corona are of the same order of size as large spot-groups, and the flare-ups, in contrast to chromospheric eruptions, last for many hours or even days. It is obviously significant that they sometimes occur, as do some magnetic storms, in the absence of visible spots on the solar disc.

Whilst the subject of coronal observation without an eclipse is under review, mention must be made of an ingenious application of television technique to astrophysical work (Skellett, 1940). In this "coronaviser" the region of the

* News has reached Great Britain recently that Edlen has identified these two infra-red lines as due to forbidden transitions in Fe XIII, the strong green line at 5303 A. to similar transitions in Fe XIV, and the red line at 6374 A. to Fe X. It appears that the identifications were made by extrapolating from the extreme ultraviolet and X-ray spectra of highly ionized atoms, and that the wavelength agreement is within the error of prediction. This attempt to identify "coronium" seems to promise more success than any of the earlier ones, but in view of the known difficulties of identifying the wide coronal lines with forbidden transitions, it would be premature to comment in detail on a subject of such first-class importance until Edlen's original paper reaches this country.

corona is spirally scanned, the circular component at 30 cycles/sec., the radial at 1 cycle/sec. The uniform sky-glare is reproduced electrically as a direct current, low-frequency components due to slow variations being filtered out, whilst variations of intensity due to prominences and coronal streamers are represented by small alternating currents capable of high amplification. Successful photographs of prominences and of some strong coronal features have already been secured with this device, and the method may in the future develop into a rival of Lyot's.

Prominence observations are a speciality of the McMath-Hulbert Observatory of the University of Michigan (Curtis, 1940), where already nearly half a million motion-picture frames have been taken with the spectroheliograph of the 50-ft. tower telescope. The films taken with the " spectroheliokinematograph " are as spectacular as Lyot's, and show a similarity to the French ones which is gratifying when one realizes that the latter are direct photographs taken with a colour filter, whereas the McMath pictures are true spectroheliograms. The data now available from these sources are already contributing towards the foundations of our theories regarding the support and motion of prominences. Our knowledge is still incomplete, though, as to the three-dimensional motion of these objects. It is difficult without such knowledge, for example, to refute suggestions that the observed motion is not material at all but is due to a travelling luminescence akin to the moving illumination produced by searchlights on terrestrial clouds. With only the x and y co-ordinates of the motion across the line of sight available, such suggestions are disturbing, though the overwhelming impression on seeing the films is that actual matter is seen in motion, especially in the cases where the form of a knot or streamer travelling along a fixed trajectory remains little changed whilst its location is completely altered. Nevertheless, the position will be clarified when the Stone spectroheliograph at the McMath-Hulbert Observatory is working fully. This instrument (McMath, 1940) gives a record of radial velocities in prominences whilst simultaneously an ordinary spectroheliograph records their cross-motions. It differs from an ordinary spectroheliograph only in the intricate motions executed by the slits. The first slit is 0·1 mm. wide and performs the function of an ordinary entrance slit; the second is 2 mm. wide and limits the length of spectrum falling on the photographic film to a strip centred on, say, the Hα line. During a first scan over the prominence and adjacent sky, the slits pause for one half-second at each of ten positions 2 mm. apart; the resultant picture is a grid of undisplaced Hα absorption lines (from the light of the sky) against which any displacements in the superposed emission line due to the prominence may be measured. After this first intermittent scan the second slit is automatically narrowed to 0·1 mm. and a continuous scan is superposed, producing a faint prominence picture which facilitates identification of the features being measured. After twenty such cycles, between each of which the image is moved by 0·1 mm. with respect to the first slit, composite photographs have evidently been obtained showing the projected form of the

prominence and the displacements of its bright line at all points with respect to the same line in absorption in the sky spectrum.

A still later paper (McMath, 1941) gives an account of the first results with this instrument. This takes the form of an exhaustive study of the structure as well as the motions of a solar prominence in three dimensions. From nearly 2000 Hα spectroheliograms of a prominence on Sept. 20, 1940, the parameters x, y, and dz/dt are measured for four easily identified knots; x and y fixing the positions of the knots in the " plane of the sky " and being obtained from spectroheliograms of the conventional type, whilst dz/dt is the radial velocity obtained as described above with the Stone instrument. For each knot a process of graphical integration then gives the true space motion as a function of t. The difficulty of arriving at the extension z along the line of sight from the observed dz/dt is ingeniously solved by making an approximate estimate of the constant of integration from subsequent observations of the prominence as a dark marking projected on the disc. From the data thus derived, a model has been constructed; and photographs of this model taken from various angles show a satisfyingly close resemblance to prominence forms frequently observed as projections on the disc or on the sky. The accelerations derived from the observed displacements and velocities are of the order of one-tenth that due to unopposed gravitation at the solar surface, though for short intervals of time accelerations exceeding the gravitational value have been observed. It is interesting to notice, in this connection, that the (distance, time) diagrams derived from these results do not show the sharp velocity changes which are the basis of Pettit's laws of prominence motions (see, for example, McMath and Pettit, 1938), and whose reality has been challenged, as regards the descending matter, by Hulme (1939). The radial velocities are of the same order as the cross-velocities, as would be expected if the observed effects are due to mass motion of the prominence atoms, as distinct from a travelling excitation. The development of this method of three-dimensional study is too recent yet for its possibilities to have been thoroughly explored, but there can be no doubt that it is destined to play a large part in elucidating one of the most puzzling problems of astrophysics.

REFERENCES

§ 1. *Telescopes and spectrographs*

ADAMS, W. S., 1941. *Astrophys. J.* **93**, 11.
BAKER, J. G., 1940 a. *Proc. Amer. Phil. Soc.* **82**, 323.
BAKER, J. G., 1940 b. *Proc. Amer. Phil. Soc.* **82**, 339.
BOWEN, I. S., 1938. *Astrophys. J.* **88**, 113.
BRACEY, R. J., 1936. *Astrophys. J.* **83**, 179.
CARTWRIGHT, C. H., 1940. *J. Opt. Soc. Amer.* **30**, 110.
HENDRIX, D. O., 1939. *Publ. Astr. Soc. Pacif.* **51**, 158.
HENDRIX, D. O. and CHRISTIE, W. H., 1939. *Scientific American*, **161**, 118.
HUMASON, M. L., 1940. *Publ. Astr. Soc. Pacif.* **52**, 145.
JONES, H. S., 1941. *Proc. Phys. Soc.* **53**, 497.

MINKOWSKI, R., 1939. *Astrophys. J.* **89**, 156.
PAYNE-GAPOSCHKIN, C. and WHIPPLE, F. L., 1940. *Proc. Nat. Acad. Sci., Wash.*, **26**, 264.
ROSS, F. E., 1940. *Astrophys. J.* **92**, 400.
WILLIAMS, R. C., 1939. *Astrophys. J.* **89**, 611.

§ 2. *Photometers and microphotometers*

ATKINSON, R. D'E., HUNTER, A. and MARTIN, E. G., 1940. *Mon. Not. R. Astr. Soc.* **100**, 196.
HALL, J. S., 1941. *Astrophys. J.* **94**, 71.
MINNAERT, M. and HOUTGAST, J., 1938. *Z. Astrophys.* **15**, 354.
MINNAERT, M., MULDERS, G. F. W. and HOUTGAST, J., 1940. *Photometric Atlas of the Solar Spectrum* (Amsterdam : D. Schnabel, Kampert and Helm).
NEWTON, H. W., 1939. *Mon. Not. R. Astr. Soc.* **99**, 463.
WILLIAMS, R. C. and HILTNER, W. A., 1940. *Publ. Obs. Univ. Michigan*, **8**, 45.

§ 3. *Coronagraphs and spectroheliographs*

CURTIS, H. D., 1940. *Pop. Astron.* **48**, 348.
HULME, H. R., 1939. *Mon. Not. R. Astr. Soc.* **99**, 634.
LYOT, B., 1939. *Mon. Not. R. Astr. Soc.* **99**, 580.
MCMATH, R. R., 1940. *Publ. Obs. Univ. Michigan*, **8**, 57.
MCMATH, R. R., 1941. *Publ. Obs. Univ. Michigan*, **8**, 123.
MCMATH, R. R. and PETTIT, E., 1938. *Astrophys. J.* **88**, 244.
SKELLETT, A. M., 1940. *Proc. Nat. Acad. Sci., Wash.*, **26**, 430.
WALDMEIER, M., 1939. *Z. Astrophys.* **19**, 21.

DISPERSION IN THE FAR INFRA-RED

By LOTTE KELLNER,
Imperial College, London

§1. INTRODUCTION

This *Report* deals with investigations, which have been undertaken within the last eleven years, into the dispersion of crystalline and liquid substances in the far infra-red ($\lambda > 10\,\mu$). This work is comparatively unknown, but is of some importance in connection with our knowledge of lattice structure, lattice forces, and the electric forces obtaining in liquids. In 1930, when the first experiments bearing on this subject were started, it was believed that everything of importance about the reflection and absorption of infra-red radiation in crystals was known. The results of Czerny's work, however, showed surprising new phenomena, results which gave a new impulse to the dynamic theory of crystal lattices.

§2. DISPERSION IN CRYSTALS

Czerny's aim (Czerny, 1930) had been to determine the refractive index and the extinction coefficient of rocksalt between 30 and $100\,\mu$, a spectral region containing the eigen frequency of NaCl. It was known from the dynamic theory of crystals (Born, 1923) that the alkali halides should exhibit one lattice vibration only in the infra-red spectrum which was to be expected in the neighbourhood of the reflection maximum ($\lambda_R = 52\,\mu$ for NaCl, $63\,\mu$ for KCl). Czerny took great pains to purify his spectrum, obtained with a wire-grating spectrometer, from all traces of wavelengths shorter than those under consideration by means of suitable filters and preliminary reflection of the incident light from residual-ray plates. He measured the transmitting and reflecting power of rocksalt, and later extended these experiments to sylvine in collaboration with Bowling Barnes (Barnes and Czerny, 1931) and calculated the optical constants from his previous observations. The reflecting power of crystals showed a very broad maximum whose centre of gravity did not coincide, however, with the calculated position of the lattice vibration; it was, therefore, essential to investigate the absorption in order to determine the eigen frequency, which should correspond to the absorption maximum. A difficulty inherent in such an investigation was, however, that of obtaining sufficiently thin plates of the crystals in question, since the crystalline absorption is already very considerable at frequencies remote from that corresponding to the lattice vibration. The transmission measurements were made at first with crystalline plates which

had been shaved down to minute thicknesses with a razor blade, but even the thinnest plate (of 8 μ thickness) still showed 90 % absorption at $\lambda = 50\,\mu$. Since it was therefore impossible to locate the maximum of absorption in this way, Czerny and Barnes obtained specimens of the requisite thickness by evaporating rocksalt and sylvine in a vacuum on to thin films of nitrocellulose which are completely transparent in the far infra-red. In this way they obtained powder layers, of 0·17 to 3·6 μ thickness, whose transmission they could investigate throughout their whole spectral range. In his first experiments Czerny had already found a subsidiary maximum of reflection at a shorter wavelength than the well-known residual rays of rocksalt. This surprising result was confirmed by the absorption measurements; two subsidiary maxima were found both in absorption and reflection for rocksalt and sylvine. Furthermore, the position of the lattice vibrations of these substances could be experimentally established for the first time. It was shown that the optical constants observed for crystalline plates agreed very well with those obtained from powdered layers in the overlapping wavelength region, so that the legitimacy of the determination of these constants from powders was proven. Czerny pointed out, besides, that interference phenomena in thin plates become only of importance on the long wavelength side of the eigen frequency λ_0, since the refractive index is below unity on the short wavelength side. The absorption maxima for $\lambda < \lambda_0$ are therefore genuine. It follows from a consideration of the relation between the transmission of an absorbing layer and its thickness that the minimum of transmission coincides with the eigen frequency for sufficiently thin plates. Czerny tried to fit the optical constants of rocksalt as observed by him into a dispersion formula. He found that he could represent the reflective power very well by using Drude's formula:

$$\left.\begin{aligned} n^2 - k^2 &= n_0^2 + \sum_{i=1}^{2} \frac{C_i(\nu_i^2 - \nu^2)}{(\nu_i^2 - \nu^2)^2 + b_i^2 \nu^2}, \\ 2nk &= \sum_{i=1}^{2} \frac{C_i b_i \nu}{(\nu_i^2 - \nu^2)^2 + b_i^2 \nu^2}, \end{aligned}\right\} \quad \ldots\ldots(1)$$

where n and k are the refractive index and extinction coefficient respectively, C_i refers to the intensity of the eigen frequency ν_i in the infra-red (the two eigen frequencies correspond to the main absorption maximum and the secondary maximum), and b_i denotes the damping term. The observed values of k agree sufficiently with the calculated values between $\lambda 30\,\mu$ and $\lambda 46\,\mu$, but below $\lambda 30\,\mu$ they are approximately two orders of magnitude smaller than those calculated. These results were confirmed by Mentzel (1934) in experiments on the absorption of NaCl, KCl and KBr which show the same deviations from Drude's theory.

Cartwright and Czerny (1933, 34) carried the determination of n and k for NaCl and KCl still further, to $\lambda 240\,\mu$, with the help of a grating spectrometer. There are no more subsidiary maxima in this region, but the extinction coefficient k, instead of decreasing towards zero, retains much higher values than equation (1) furnishes.

When the existence of secondary reflection maxima in rocksalt and sylvine had been established, other substances were investigated, and in all instances subsidiary maxima were found. Barnes (1932) extended the experiments on the absorption of powders to seventeen alkali halides. In nearly all cases, secondary absorption bands were observed. If the atomic weight of the metal ions is plotted as abscissa against the eigen frequency as ordinate, smooth curves are obtained.

The reflective power of artificially grown LiF and NaF crystals was measured by Korth (1932); both substances show a narrow, sharp, subsidiary maximum on the short wavelength side of the residual rays. The case of LiF is remarkable on account of the fact that the subsidiary maximum reaches an intensity of 80 %, while normally this value is between 40 and 50 %. The reflective power of LiF is slightly increased at low temperatures (87° K.) to practically 100 % in the main maximum, but the separation between the two maxima is not improved.

Table 1. Wavelengths of absorption and reflection maxima of face-centred cubic crystals

Substance	Fundamental lattice vibration	Secondary vibrations	Residual rays	
			Main maximum	Subsidiary maxima
	μ	μ	μ	μ
Lithium fluoride, LiF	32·6	19 ; 26	26	16·8
Sodium fluoride, NaF	40·6	34	35·8	24
Thallium fluoride, TlF	67·5		48·6	
Sodium chloride, NaCl	61·1	34 ; 40·5 ; 51	52·0	38 ; 48
Potassium chloride, KCl	70·7	33 ; 41 ; 42·5 47 ; 60	63·4	
Rubidium chloride, RbCl	84·8	77	73·8	
Caesium chloride, CsCl	102·0	97		
Thallium chloride, TlCl	117·0	112	91·9	
Sodium bromide, NaBr	74·7	67		
Potassium bromide, KBr	88·3	35 ; 83	81·5	
Rubidium bromide, RbBr	114·0	108		
Caesium bromide, CsBr	134·0	127		
Thallium bromide, TlBr	beyond range of spectrometer		117	
Sodium iodide, NaI	85·5	80		
Potassium iodide, KI	102·0	97	94	
Rubidium iodide, RbI	129·5	125		
Thallium iodide, TlI	beyond range of spectrometer		151·8	
Magnesium oxide, MgO	17·3	6·7 ; 14·9 ; 25	15·4	14·8

Hohls (1937) completed these investigations by absorption and reflection measurements down to $\lambda 55\,\mu$. He used artificial crystals of LiF and NaF and obtained the required thickness by etching thick plates and afterwards heating them to 300° C. The subsidiary bands were observed in absorption in addition to the lattice vibrations. From the observed transmission and reflection data Hohls calculated the refractive indices and extinction coefficients of LiF and NaF in the same way as Czerny.

In table 1 the experimental data published so far on the lattice vibrations of face-centred cubic crystals have been collected. Wherever possible, the position of the absorption maxima, coinciding with the true eigen frequency, has been given, together with that of the residual rays (reflection maxima). The substances concerned are all alkali halides with the exception of MgO.

Contradictory results have been published about the absorption of MgO, a face-centred cubic crystal of the NaCl type. According to Fock (1934), who burnt Mg ribbon and collected the MgO fumes on nitrocellulose films in layers from a few μ to 0·2 mm. thickness, these fine crystalline deposits showed a strong absorption maximum at $\lambda 17{\cdot}3\,\mu$ flanked on both sides by minor maxima at $\lambda 14{\cdot}9\,\mu$ and $\lambda 25\,\mu$. A weak absorption band was also found at $\lambda 6{\cdot}7\,\mu$. From $\lambda 50\,\mu$ to $\lambda 200\,\mu$, MgO has a transmission of 90 to 100%. The measurements were carried out at room temperature and 193° K; no temperature effects out of the ordinary occurred.

Barnes, Brattain and Seitz (1935) obtained quite extraordinary results in their experiments on the absorption and reflection of MgO. They used for their investigations thin crystalline plates cleaved from large, artificially grown crystals as well as crystalline layers of MgO evaporated *in vacuo* on to nitrocellulose films and the fumes of burnt Mg ribbon deposited on rocksalt plates. All three samples gave consistent results in their absorption experiments. They did not find an absorption band between $\lambda 6\,\mu$ and $\lambda 7\,\mu$, but observed a great number of narrow, weak, absorption bands from $\lambda 6\,\mu$ onwards, superimposed on the main absorption maximum at $\lambda 15{\cdot}4\,\mu$. Since these investigators used a rocksalt prism whose range extended only as far as $\lambda 16\,\mu$, they were not able to observe the principal maximum at $\lambda 17{\cdot}4\,\mu$ obtained by Fock. The strange feature of their MgO spectrum is the occurrence of this great number of narrow secondary maxima. A cleaved plate of 0·075 mm. thickness shows 40 bands of that kind between $\lambda 6\,\mu$ and $\lambda 15{\cdot}5\,\mu$, i.e. they appear at a mean distance of $\lambda 0{\cdot}24\,\mu$, corresponding to an interval of about 14 cm.$^{-1}$. They cannot be due to interference effects since, as Czerny has pointed out, interference is unimportant on the short wavelength side of an eigen frequency. On the other hand, it follows, from the curves published by Barnes, Brattain and Seitz, that this secondary structure of the MgO spectrum decreases in complexity with increasing thickness of the crystal. Thus, for a crystal of 3·05 mm. thickness, only one band is found per wavelength interval of $1\,\mu$, but in a thicker plate the same or even a greater number of bands of stronger intensity should be observed if they originate from selective

absorption, as Barnes and Czerny have shown for NaCl and KCl. The genuineness of this secondary structure is, therefore, doubtful in face of Fock's results and the peculiar behaviour of the spectra obtained by Barnes, Brattain and Seitz, until more experimental evidence is forthcoming. In reflection two maxima of practically equal intensity were observed at $\lambda 14{\cdot}8\,\mu$ (62 %) and $\lambda 15{\cdot}4\,\mu$ (65 %).

The investigations considered hitherto are of twofold importance: firstly, they led to the discovery of the secondary structure of the absorption and reflection spectra of crystals, and secondly, they provided the first direct determinations of the infra-red eigen frequencies, which up till then could only be indirectly worked out from the refractive indices.

Parodi continued this latter work for substances other than alkali halides. His experiments were carried out with the help of twelve residual-ray groups covering the wavelength range $\lambda 20{\cdot}75\,\mu$ to $\lambda 63\,\mu$. He measured the absorption of crystalline powders deposited on paraffin plates and also the reflection of crystals and fused polished plates of some of the substances under consideration. His experiments are concerned with borates (Parodi, 1937), and with oxides, fluorides and some copper halides (Parodi, 1938). His data are collected in table 2.

The experiments carried out by Czerny and his collaborators had shown that the lattice theory did not explain the finer structure of the far infra-red spectrum of crystals since, according to Born (1923), cubic crystals should exhibit only one active infra-red frequency, in which one type of ions oscillates against the other. Furthermore, modifications of the theory of dispersion were required since the behaviour of the extinction coefficient outside the region of anomalous dispersion is at variance with the predictions of Drude's theory. Born and Blackman (1933) showed, following the publication of Czerny's and Korth's results, that the subsidiary maxima in the absorption and reflection spectrum of crystals resulted from the anharmonicity of the lattice forces. If we consider the case of a linear chain with anharmonic forces between neighbouring ions of opposite charge and respective masses m and M, it appears that the fundamental vibration becomes coupled to certain combinations between other normal vibrations which are inactive for harmonic forces. Consequently, an energy transfer between the fundamental and the combination frequencies becomes possible, which results in the damping of the main vibration. Two combinations of that kind are permitted for the linear chain,

$$\left.\begin{aligned} \nu_1+\nu_2 &= \nu_0(m^{\frac{1}{2}}+M^{\frac{1}{2}})/(m+M)^{\frac{1}{2}}, \\ \nu_1-\nu_2 &= \nu_0(m^{\frac{1}{2}}-M^{\frac{1}{2}})/(m+M)^{\frac{1}{2}}, \end{aligned}\right\} \qquad \ldots\ldots(2)$$

if ν_0 is the fundamental lattice vibration. The summation tone should occur at approximately $\nu_0\sqrt{2}$ on the short wavelength side of ν_0 and the difference tone on the long wavelength side in the extreme infra-red. Blackman (1933, 36) extended this theoretical treatment to the three-dimensional lattice of the NaCl

Table 2. Wavelengths of absorption and reflection maxima of borates

Substance	Absorption maxima		Reflection maxima			
	μ	μ	μ	μ	μ	μ
Calcium borate, $CaB_4O_7 . 4H_2O$	27·8 30·8	40	52	31	42	47·5
Potassium borate, KBO_2	28	39	52	29	35	48
Manganese borate, $Mn(BO_2)_2$	30	43	52			
Cupric metaborate, $Cu(BO_2)_2$	30	38	52			
Lead borate, $Pb(BO_2)_2 . H_2O$	30·8	43	52	33	44	47

Wavelengths of absorption maxima of oxides

Substance	Absorption maxima				
	μ	μ	μ	μ	μ
Calcium oxide, CaO		27·4			
Cobalt monoxide, CoO		28		38·5	
Manganese monoxide, MnO			33		
Cupric oxide, CuO			31	38	
Zinc oxide, ZnO	22·4	30			
Strontium oxide, SrO					47
Cadmium oxide, CdO				40	51·5
Barium oxide, BaO					53
Lead monoxide, PbO			30·8		
Titanium dioxide, TiO_2			30·5	41	
Manganese dioxide, MnO_2		29		40·7	50
Molybdenum dioxide, MoO_2		27		40	51·5
Lead peroxide, PbO_2			30·5	42	53
Alumina, Al_2O_3			30·5	41	51·5
Manganese sesquioxide, Mn_2O_3		28·5	33·6	42·8	53·5
Nickel sesquioxide, Ni_2O_3		28·5	33	42·6	52
Antimony trioxide, Sb_2O_3	23	28			56
Zinc sulphide, ZnS			33		
Cuprous chloride, CuCl					53
Cuprous bromide, CuBr					57
Cuprous iodide, CuI					58

Wavelengths of absorption and reflection maxima of fluorides

Substance	Absorption maximum	Reflection maximum
	μ	μ
Calcium fluoride, CaF	51·5	31
Strontium fluoride, SrF	69	40·6
Barium fluoride, BaF	73	43

The great interval between the lattice vibration and the reflection maximum in the case of the fluorides, in the above table, is specially to be noted.

type where the forces between mass points at distances a, $a\sqrt{2}$ and $a\sqrt{3}$ (a is the lattice constant) have to be taken into account, involving three force constants.* Correspondingly, the optical and acoustical frequency branches of the linear chain split into three parts each, and a great number of combination tones becomes possible.

The formulae from which these vibrations can be calculated contain, in addition to the masses of the ions, the three force constants which are connected with the elastic constants c_{11} and c_{12}. These combination tones again consist of summation tones forming one or more groups on the short wavelength side of the fundamental vibration, and difference tones in the extreme infra-red.

Barnes, Brattain and Seitz developed a quantum-mechanical theory of the three-dimensional lattice of the NaCl type, using group-theoretical methods and an anharmonic potential. An extremely complicated secondary structure for the spectrum of MgO results from their assumption that the density maxima of the normal vibrations coincide with the extreme points of the frequency spectrum, which latter are determined by the authors. Since this assumption probably conflicts with the actual conditions in a lattice, this analysis would have to be further investigated. Barnes, Brattain and Seitz have not attempted

Table 3. Wavelengths of secondary maxima

Substance	Summation tones			Difference tones		
	Observed	Calculated		Observed	Calculated	
	μ	μ	μ	μ	μ	μ
Sodium chloride NaCl	34		35·0	~200		222
			36·6			333
			38·6	λ>200	390	382
	40·5		39·1			562
		43·3	41·4			698
	51		46·7			825
Potassium chloride, KCl	33			200		
	41		45·5	λ>200		763
	42·5		45·6			1122
			46·3			1350
	47		46·9			1760
			47·6		2020	2200
		49·9	49·5			3300
	60					
Magnesium oxide, MgO	6·7					
		12·3	12·5			178
			13·0		119	125
			13·8			79
			14·0			418
	14·9		15·0			143
			15·5			218
	25					

* See also the *Report* by Blackman in this volume, p. 11—[Editor].

to apply their theoretical results to their experimental data, as they are not certain that the $\lambda 15{\cdot}4\,\mu$ band corresponds to the fundamental lattice vibration (it has been explained above that it does not represent the eigen frequency), and as the resolving power of their spectrometer is not sufficient between $\lambda 5\,\mu$ and $\lambda 10\,\mu$ to warrant the assumption that their observations are final.

In table 3 the observed subsidiary absorption maxima of NaCl, KCl and MgO are compared with the calculated combination tones. In columns 2 and 5 are given the experimental data, in columns 4 and 7 the results of the three-dimensional theory, and in columns 3 and 6 the frequencies following from the simple formulae (2) for the linear chain.

In the case of MgO, λ_0 has been taken to be $17{\cdot}3\,\mu$. It will be seen that the values on the short wavelength side calculated from the three-dimensional lattice theory for NaCl form three groups corresponding to the three observed subsidiary maxima; for KCl, however, the agreement is less satisfactory, the theory failing to yield the $\lambda 60\,\mu$ band. Although the one-dimensional theory gives a secondary band near the values observed for NaCl and KCl, the corresponding value for MgO appears at a much shorter wavelength. The three-dimensional theory, however, furnishes better agreement with the experiments. On the other hand, Fock could not detect any subsidiary absorption maxima between $\lambda 50\,\mu$ and $\lambda 200\,\mu$, while the theory demands absorption in this region. It follows from a comparison between observed and calculated data that the assumption of anharmonic forces in the lattice serves to interpret the features of the spectrum qualitatively.

In his treatment of the vibrations of a linear chain, Blackman (1939) investigated how the damping of the oscillations depended on the positions of the subsidiary maxima and obtained an explicit dispersion formula. He showed that the damping is of the form

$$\frac{\beta k T(m+M)^3}{\pi^4(mM)^{\frac{1}{2}}} \cdot \frac{\nu}{\nu_0^4\sqrt{(\nu_a^2-\nu^2)(\nu^2-\nu_b^2)}} \quad \text{for } \nu_a<\nu<\nu_b,$$

where β is the force constant determining the anharmonicity of the lattice potential, k Boltzmann's constant, T the absolute temperature, ν_0 the fundamental vibration and $\nu_a=\nu_1+\nu_2$, $\nu_b=\nu_1-\nu_2$ given by (2). The damping is, therefore, frequency-dependent and not constant as in Drude's theory, and leads to a dispersion formula of the following form:

$$\left.\begin{aligned} n^2-k^2 &= n_0^2+\frac{C_1(\nu_0^2-\nu^2)(\nu_a^2-\nu^2)(\nu^2-\nu_b^2)}{(\nu_0^2-\nu^2)^2(\nu_a^2-\nu^2)(\nu^2-\nu_b^2)+\gamma^2\nu^4},\\ 2nk &= \frac{C_2\gamma\nu[(\nu_a^2-\nu^2)(\nu^2-\nu_b^2)]^{\frac{1}{2}}}{(\nu_0^2-\nu^2)^2(\nu_a^2-\nu^2)(\nu^2-\nu_b^2)+\gamma^2\nu^4}, \end{aligned}\right\} \quad \ldots\ldots(3)$$

where n, k, n_0 have the same meaning as in (1), C_1 and C_2 determine the intensity of the maxima of n and k, and γ contains the constant terms in the damping

factor. Formula (3) is valid within the limits $\nu_a < \nu < \nu_b$; outside these limiting values the damping becomes zero, i.e. $k = 0$ and

$$n^2 = n_0^2 + C_1/(\nu_0^2 - \nu^2).$$

It will be seen that the sum over the eigen frequencies in (1) has been replaced by a product over the same expressions and that n and k will assume extreme values in the neighbourhood of ν_0, ν_a and ν_b. An attempt is made here to calculate n and k from (3) by putting $\lambda_0 = 61{\cdot}1\,\mu$, $\lambda_a = 40{\cdot}5\,\mu$ (these values are taken from Czerny's work) and $\lambda_b = 380\,\mu$ (calculated from the one-dimensional theory). C_1 and C_2 and γ were chosen so as to fit the observed values of n and k for $\lambda = 43\,\mu$ and $46\,\mu$; no attempt has been made to smooth out the curves by taking more values of n and k into account. The results are set out in table 4.

In the columns of results " calculated according to Drude " use has been made of Czerny's calculations of n and k by means of equation (1) with two resonance positions at $\lambda_1 = 61{\cdot}1\,\mu$ and $\lambda_2 = 40{\cdot}5\,\mu$.

It follows from table 4 that equation (3) gives the steep drop in k, observed in the case of short waves—a region in which Drude's dispersion theory fails completely for the computation of the extinction coefficient. Furthermore, it leads to comparatively high absorption values on the long wavelength side of the fundamental vibration which is in agreement with experiment. Here, Drude's theory demands a sharper decrease of the extinction coefficient than has been observed. Equation (3) agrees therefore with the general features of the absorption spectrum much more satisfactorily than did the older dispersion

Table 4. Refractive index and extinction coefficient of NaCl

λ in μ	Refractive index n			Extinction coefficient k.		
	Observed	Calculated according to		Observed	Calculated according to	
		Drude	Blackman		Drude	Blackman
10	1·49	1·495	1·48	$\sim 10^{-7}$	$\sim 10^{-4}$	0
17	1·42	1·4	1·40	4×10^{-4}	4×10^{-3}	0
30	not observed	1·1	1·26	0·04	0·04	0
40·5	0·52	0·52	0	0·68	0·68	1·53
46	0·20	0·18	0·54	1·17	1·1	1·27
50	not observed		0·46	not observed		1·75
54	,, ,,		0·75	,, ,,		2·09
60	,, ,,		1·92	,, ,,		1·80
61·0	,, ,,	6·78	2·24	,, ,,	6·3	1·64
75	3·42	3·85	3·46	0·133	0·2	0·64
100	2·70	2·86	2·98	0·083	0·06	0·18
120	2·51		2·82	0·080	0·04	0·14
150	2·51		2·72	0·070	0·015	0·11
200	2·51		2·66	0·056	0·01	0·11
240	2·50		2·63	0·04	0·008	0·11

theory ; its drawback is that in the present form it furnishes a reflective power of 100 % for the subsidiary maxima ν_a and ν_b, while for the main residual rays at $\lambda 52\,\mu$ only 65 % is obtained.

It is well known that the maximum reflection occurs for all crystals at a shorter wavelength than the eigen frequency if plates of ordinary thickness are used. But, as Blackman (1934) has pointed out, the position of the reflection maximum as well as its intensity depends on the thickness of the material. With decreasing thickness of the plate the reflective power decreases, and the maximum shifts towards longer wavelengths, i.e. towards the eigen frequency. In the case of NaCl it has been shown that the reflection maximum for a crystal of $0{\cdot}2\,\mu$ thickness coincides with the absorption maximum, and its intensity has dropped from 90 % to 24 %. Besides, the secondary maxima have completely disappeared.

§ 3. DISPERSION OF LIQUIDS

Using the methods developed by Czerny and his collaborators, Cartwright and Errera (1935, 36) determined the optical constants of a number of liquids between $\lambda\,50\,\mu$ and $\lambda\,150\,\mu$. They covered this wavelength region with a series of residual rays and investigated thirteen polar liquids (including some alcohols, ethers and light and heavy water) and seven non-polar liquids (such as hexane, CCl_4, benzene, carbon disulphide, etc.). Several substances showed absorption bands in the far infra-red which remained undisturbed on dilution with non-polar liquids of low absorption, with the exception of the $\lambda 60\,\mu$ band of H_2O and D_2O which disappeared. It is to be concluded that the undisturbed bands are caused by intramolecular vibrations; the water band, on the other hand, is connected with the intermolecular forces in the liquid.

The main object of these authors was to determine the atomic polarization of these substances from the observed optical constants n and k. The total polarization P of a liquid is the sum of three terms P_e, P_a and P_p, which are, respectively, the polarization caused by the atoms, the electrons and the dipoles in the liquid. In the case of non-polar liquids, P_p is zero. The total polarization is given by the static dielectric constant. $P_e + P_a$ is calculated from the infra-red refractive index, molecular weight and density of the substance in the same way as P_e is calculated from the refractive index in the case of visible light. P_p is taken from the measurements of the relation between dipole moment and temperature. Cartwright and Errera use Debye and Sack's theory of the relation of the relaxation times of polarizable molecules under the influence of light to work out these data from the observed refractive indices and extinction coefficients. The atomic polarizations determined by them agree much better with the values for gases than with the values obtained for liquids by previous more indirect methods. From the relaxation times the molecular radii can be calculated using the assumptions that Stokes's law is valid and that the molecules are spherical in shape. They come out much too small as compared with other

determinations, a not unexpected result, since the molecules in an organic liquid approximate more closely to an elliptical configuration.

Cartwright (1936 a, 36 b) continued this work in the same spectral region with electrolytes. He investigated the absorption and reflection of HCl, LiCl, NaCl, KCl, KBr, LiI, KI, ZnI_2, H_2SO_4, $MgSO_4$ and KOH for different concentrations in aqueous solution. On applying Maxwell's equations to this problem, a characteristic wavelength, λ_0, of the incident light is found below which the inertia and viscosity of the ions are so great that they can no longer follow the alternations of the electromagnetic waves. Above λ_0 the ions perform a periodic translational motion which decreases in intensity with increasing frequency of radiation; this is a consequence of the inertia of the ions and their friction against neighbouring molecules. It is assumed here that the binding forces between the ions and the surrounding water molecules are negligible. If Stokes's law holds,

$$\lambda_0 = 1{\cdot}65 M/\eta r,$$

where M is the mass of the ion and η the specific viscosity of the solution, r the radius in A. For KCl and KI in solution, the corresponding calculated characteristic wavelengths are

$$\lambda_0^{K} = 48\,\mu,\ \lambda_0^{Cl} = 32\,\mu,\ \lambda_0^{I} = 95\,\mu.$$

The assumption that the ions are free to move independently of the water molecules surrounding them is compatible with the experimental data on the far infra-red absorption in the cases of KCl and KI; these two electrolytes moreover, have, in aqueous solution, practically the same viscosity as pure water. The results for LiCl and $MgSO_4$ on the other hand, indicate that here the ions move together with the water molecules, i.e., the binding forces are no longer negligible. The facts that these two solutions are considerably more viscous than water and have abnormally small electrical conductivities, point to the same conclusion.

In the case of H_2SO_4 it is shown that the pure acid absorbs and reflects less than H_2O, a result which is predicted by Debye's dipole theory. H_2SO_4, concentrated or dilute, has an absorption band at $\lambda\,60\,\mu$ which Cartwright ascribes to the SO_4 radical. The reflection and absorption of the dilute acid are considerably above that of either the pure acid or pure water, which is compatible with the presence of ions in the dilute solution. Cartwright has studied the optical properties of H_2SO_4 solutions for different concentrations and concludes that a dilution with less than 5 % H_2O leads to the formation of OH_4^{++} and SO_4^{--} ions; further dilution with water furnishes H^- ions and H_2SO_5 molecules, and, at even greater dilution, $H_3^+ SO_4^{--}$ ions.

Though some of these conclusions may have to be modified in the light of less simple assumptions, Cartwright and Errera's work will still be of great

importance, since it has revealed completely new aspects in the spectroscopy of the far infra-red and has introduced somewhat more direct methods for the determination of the polarization and internal structure of liquids and electrolytes.

REFERENCES

BARNES, R. B., 1932. *Z. Phys.* **75**, 723.
BARNES, R. B., BRATTAIN, R. R. and SEITZ, F., 1935. *Phys. Rev.* **48**, 582.
BARNES, R. B. and CZERNY, M., 1931. *Z. Phys.* **72**, 447.
BLACKMAN, M., 1933. *Z. Phys.* **86**, 421.
BLACKMAN, M., 1934. *Phil. Mag.* **18**, 262.
BLACKMAN, M., 1936. *Philos. Trans.* **236**, 103.
BLACKMAN, M., 1939. *Z. Phys.* **112**, 256.
BORN, M., 1923. *Atomtheorie des festen Zustandes* (Berlin).
BORN, M. and BLACKMAN, M., 1933. *Z. Phys.* **82**, 551.
CARTWRIGHT, C. H., 1936 a. *Phys. Rev.* **49**, 101.
CARTWRIGHT, C. H., 1936 b. *J. Chem. Phys.* **4**, 413.
CARTWRIGHT, C. H. and CZERNY, M., 1933. *Z. Phys.* **85**, 269.
CARTWRIGHT, C. H. and CZERNY, M., 1934. *Z. Phys.* **90**, 54.
CARTWRIGHT, C. H. and ERRERA, J., 1935. *Acta physicochimica, U.R.S.S.*, **3**, 649.
CARTWRIGHT, C. H. and ERRERA, J., 1936. *Proc. Roy. Soc.* A, **154**, 138.
CZERNY, M., 1930. *Z. Phys.* **65**, 600.
FOCK, J., 1934. *Z. Phys.* **90**, 44.
HOHLS, H. W., 1937. *Ann. Phys., Lpz.*, **29**, 433.
KORTH, K., 1932. *Gött. Nachr., math.-naturwiss. Klasse*, 576.
MENTZEL, A., 1934. *Z. Phys.* **88**, 178.
PARODI, M., 1937. *C.R. Acad. Sci., Paris*, **204**, 1111 and 1636 ; **205**, 906 and 1224.
PARODI, M., 1938. *C.R. Acad. Sci., Paris*, **206**, 1717.

THE M.I.T. WAVELENGTH PROJECT

By GEORGE R. HARRISON,
Massachusetts Institute of Technology, Cambridge, Mass.

§ 1. INTRODUCTION

PHYSICISTS have revealed in recent years an increasing tendency to consider the major portion of the spectroscopist's work done, in so far as the assembly of the raw materials of atomic spectra is concerned. Yet anyone who seeks wavelength data for the classification of spectra and the determination of energy levels in atoms will realize that great gaps still exist in our knowledge of spectrum lines. Data regarding approximately one million atomic and ionic lines will probably be needed in order to give astronomers, physicists, and chemists all of the material of this sort which they can effectively use. A thorough search of the literature reveals that not more than 280,000 spectrum lines have thus far been measured and assigned to parent atoms. The job of providing the raw material of line spectroscopy must be considered as little more than one-fourth complete.

Two types of wavelength catalogues are needed. One of these should list the wavelengths and relative intensities of all the spectrum lines to be expected from each atom in each stage of ionization. The other should list in regular order the wavelengths and intensities of all known spectrum lines, giving after each the atom or ion responsible for it. To learn which lines are to be expected from a given element, one would then consult a catalogue of the first type. To identify the atom responsible for any particular line whose wavelength has been determined, one would consult a catalogue of the second type.

The most extensive catalogue of the first type is the monumental *Handbuch der Spectroscopie* of H. Kayser,* which lists about 120,000 lines, though for some 40 elements it contains no data determined since 1912. The most extensive catalogue of the second type is the *M.I.T. Wavelength Tables*,† which contain the 109,000 strongest lines of the elements.

During the past ten years the measurement of wavelengths has lost much of its character of original research and has taken on certain aspects of dull routine. The discovery and measurement of further spectrum lines has become progressively more difficult and less interesting. A few thousand lines account for nine-tenths of the light usually emitted by atoms, and naturally these stronger lines were observed and measured first. Yet the further measurement of wavelengths and classification of spectrum lines is of primary importance to science—

* H. Kayser, and Kayser and H. Konen (Leipzig: Hirzel, 1910-34).
† G. R. Harrison, Editor (New York: John Wiley and Sons, 1939).

the astrophysicist and the chemist will probably in the future obtain even more important information from the data of spectroscopy than they have in the past.

At present our quantitative knowledge of the extra-nuclear structure of the atom may be considered as reasonably complete for only a few atoms. We have a number of semi-quantitative ideas regarding the structure of the remaining eighty-odd atoms, but in very few cases is the exact knowledge available which is required to explain in detail such matters as chemical bonding. The whole modern science of theoretical chemistry is based largely on spectroscopic data, and its further development is handicapped by the paucity of such data. This fact has been obscured by the difficulties presented by the complicated methods of calculation still needed to solve problems in theoretical chemistry.

The spectroscopist of 1890 was likely to discover an important generalization as the result of measuring only a few dozen spectrum lines, yet today painstaking work on a thousand lines may yield little of obvious and immediate importance. The first spectrum of sodium could be analysed when the wavelengths of two or three hundred lines had been measured to ± 1 A. To analyse the various spectra of neodymium will require that the wavelengths of several thousand lines be known to within a few thousandths of an angstrom.

Considerations of this sort made apparent the desirability of developing new methods of assembling the raw materials of spectroscopy. Therefore, when the Spectroscopy Laboratory of M.I.T. was installed in a new building ten years ago, attention was directed to devising wholesale methods of wavelength measurement and reduction of spectrum lines. The present *Report* contains a summary of the apparatus developed and the results obtained in a project involving the measurement of spectrum lines, the classification of spectra and the procurement of other new data valuable in the analysis of atomic spectra.

§ 2. THE DEVELOPMENT OF NEW APPARATUS

Recent improvements in concave diffraction gratings, involving particularly the use of aluminized surfaces of high reflectivity and controlled groove shape whereby most of the light is thrown into one order, now make possible the spectrum photography of sources having numerous faint lines. This laboratory is equipped with five large diffraction gratings, some having 15,000 and others 30,000 lines per inch, ruled by Professor R. W. Wood at Johns Hopkins University. These are: (1) a six-inch grating of 165,000 lines in a 10-metre modified Paschen mounting; (2) a six-inch grating of 170,000 lines, in a 10-metre modified Rowland mounting; (3) a six-inch grating of 120,000 lines and 10-metre radius, in a stigmatic Wadsworth mounting, used mostly in the range 1950 to 2500 A.; (4) a seven-inch 10-metre grating of 90,000 lines, in a stigmatic Wadsworth mounting, used mostly for the range 5,000 to 10,000 A; (5) a six-inch speculum-metal grating of 180,000 lines and 21 ft. radius, coated with aluminium and mounted in vacuum at normal incidence, for the range 500 to 2000 A. The dispersions thus obtained range from 0·33 A. per mm. to 3·3 A. per mm., the latter value being used only in the infra-red.

With these spectrographs, whose detailed specifications have been described elsewhere (Harrison, 1933 a, 33 c; Harrison and McNally, 1940), more than 10,000 2 in. × 20 in. spectrograms have been made in the course of the wavelength projects now under discussion. These plates contain photographs of some millions of spectrum lines, and the labour of measuring and reducing this data would have been prohibitive were it not for the introduction of labour-saving devices. To aid in the reduction of data three machines were developed: (1) the automatic comparator (Harrison, 1935, 38 a; Harrison and Molnar, 1940), capable of measuring plates up to 25 inches long, which records wavelengths on motion-picture film in angstrom units to seven figures, at the rate of more than a thousand a minute; (2) the spectroscopic interval sorter (Harrison, 1932, 33 b), which in effect automatically subtracts each seven-figure wavenumber in a list from every other wavenumber within a desired frequency interval, at the rate of 100,000 subtractions a minute; at the same time it sorts the differences into an array such that wavenumber differences which occur more often than expected by chance become immediately obvious; (3) the interval recorder (Harrison, 1938 b), which, using results obtained from the interval sorter, determines all pairs of lines in a spectrum list which possess the constant differences which the interval sorter has found to be of importance.

The two last-named devices have been found useful in the application of the Ritz combination principle to the analysis of spectra in which wavelengths of high precision have been made available with the automatic comparator.

The work of the project has also been greatly facilitated by the availability of a new type of electro-magnet developed by Professor Francis Bitter (Bitter, 1939; Bitter and Harrison, 1940), in which fields up to 100,000 oersteds can be produced for Zeeman-effect work.

These various devices have closely related functions, since wavelength determination, line classification, and Zeeman-effect studies are very closely interdependent. Without exact wavelength measurements a complex-term analysis is difficult to make even if excellent Zeeman observations are available for many lines. On the other hand, wavelength measurements cannot be made with high precision without satisfactory working standards of wavelength, and the standards needed for high dispersion can best be determined nowadays by use of the combination principle, which can be applied only in cases where the term analysis is known. Yet, in a very complex spectrum, even the highest wavelength precision is insufficient to make a term analysis possible merely by application of the combination principle, and Zeeman-effect studies are almost essential. The more complex the spectrum, the smaller is likely to be the proportion of lines for which satisfactory Zeeman resolution can be obtained; and Zeeman measurements themselves are wavelength determinations and hence involve similar technique to that used for determining other wavelengths. Our experience has emphasized the great desirability of having in the same laboratory facilities for wavelength measurements, Zeeman-effect measurements, and term classification.

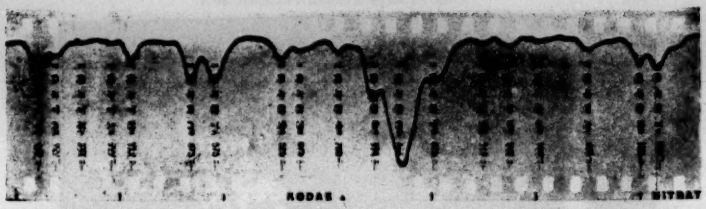


Figure 3. Section of a typical record from the automatic comparator, showing the density trace of a small portion of a spectrogram, with wavelengths recorded for each line.

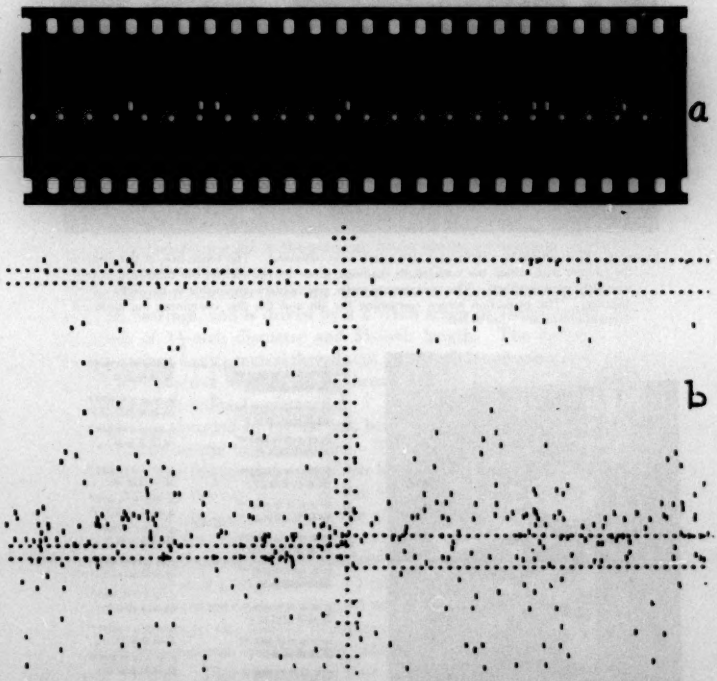


Figure 6. *a*. Photograph of a short section of the punched tape. Rectangular holes represent the wavenumbers of spectrum lines. Sets of round holes are punched only at intervals of a few hundred wavenumbers to produce fiducial marks so that the resulting graph can be read easily. *b*. Small portion of a record obtained from the spectroscopic interval sorter. Wavenumber difference is plotted horizontally and wavenumber vertically.

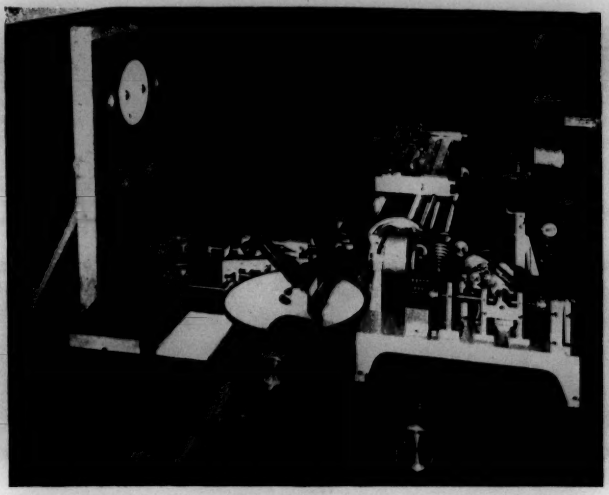


Figure 1. The automatic comparator with glass cover removed. The black case at the right is the camera with which the wavelength readings, given by the dial on the right, are instantaneously photographed. The comparator screw and carriage are visible in the right centre distance. The projection screen containing the slit and the case containing the electrical "maximum-picker" are at left.

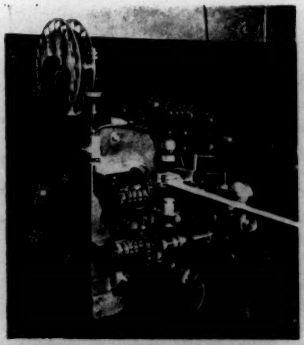


Figure 8.

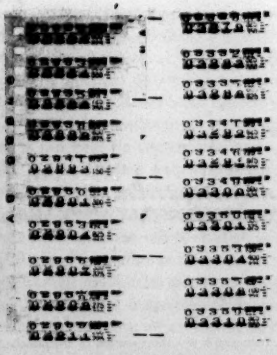


Figure 9.

Figure 8. The spectroscopic interval recorder, showing the wavenumber dials.

Figure 9. Short sections of two records from the spectroscopic interval recorder.

To face page 215

Though the scientific personnel available for the projects under discussion have been somewhat limited in numbers, their efforts have been very effectively supplemented by the labours of clerks furnished by the Works Projects Administration. Various problems arose, naturally, because these workers were unfamiliar with the background of spectroscopy, but the use of the automatic equipment to be described helped greatly in reducing this handicap.

§ 3. THE AUTOMATIC COMPARATOR

The automatic comparator contains three improvements over ordinary measuring engines: automatic determination of the centres of spectrum lines with high precision and without pause; automatic transformation of distances along the plate into wavelengths in angstroms with seven-figure precision; and automatic recording on film of the resulting wavelengths, together with an intensity trace of the line contours. The time needed to measure a 20-inch plate is 120 seconds, and the machine is overloaded only if the number of lines on a single plate is greater than four thousand.

A general view of the automatic comparator is shown in figure 1. In this view the glass cage surrounding the instrument has been removed so that the arrangement of the various parts may be more clearly seen. From one to three spectrum plates are clamped to the carriage shown in the rear centre of figure 1, and one of these is racked into place to be moved at constant speed across a beam of light, which throws on a slit in the circular white screen an image of that portion of the plate passing across it.

The comparator carriage rolls on three sets of wheels mounted on very accurate ball bearings, and is driven by a 27-inch screw of 1-mm. pitch, cut on a steel blank of $1\frac{1}{2}$-inch diameter and 35-inch length. The half-nut which pushes the carriage has its central threads cut away both longitudinally and transversely, to give positive bearing on all thread surfaces and good dust-clearing properties, yet low frictional resistance.

The screw is mounted in special ball bearings at both ends, to reduce the necessary driving torque to a minimum. The gears used for automatic computation add to the frictional drag, yet the use of ball bearings throughout the instrument reduces driving torque to less than that required for an ordinary comparator. The operator can move the plate in either direction by turning the handle in the lower right-hand corner of figure 1 or by operating an electric drive providing four speeds forward and four in reverse.

The lapping of such a long screw to ± 1 micron error is practically impossible, so a new type of corrector bar was designed which would remove both cumulative and periodic errors to the desired accuracy. Both types of correction are cut into a three-dimensional rotating cam which has been described elsewhere (Harrison, 1938 a). This cam can be seen to the right of and just below the comparator carriage, running parallel to the screw and ways.

An automobile headlamp operated on a storage battery is mounted below the

comparator carriage. Light from this lamp is focused by a condenser lens, through the plate being measured, on to a micro-tessar lens which produces an enlarged image of the plate on the screen some 6 ft. distant. Three projection lenses are provided in a rotating nosepiece, giving magnifications of from 10 to 120 diameters, 30 being the value most commonly used. The slit in the screen can be varied in length from 0 to 3 inches, and in width from 0 to 0·5 inch, and is normally set at the minimum width consistent with the graininess of the emulsion used.

A Farnsworth electron-multiplier tube of the electrostatic-focusing type is used behind the slit to measure the light passing through it. This has 11 stages, which give an overall current amplification of several million. The output of the multiplier is fed to an electrical network, which delays uniformly by ·013 second any signal which can be resolved into frequencies lying between 0 and 200 cycles. The attentuation of this network is practically uniform over the frequency range desired. The difference between the output and an equal fraction of the input voltage of the network is applied to a d.c. amplifier. This voltage difference falls to zero ·0065 second after the centre of a spectrum line has been imaged symmetrically on the slit, when the line is moved at constant velocity across the slit at the speed used.

The time delay of ·013 second corresponds to a distance along the plate of 0·025 mm. when the plate is being scanned uniformly at 4 mm. per second. When the voltage difference mentioned above falls to zero, a " zero-picker " circuit causes a stroboscopic lamp to flash by discharging a condenser through it. The mechanism of production of the pulse which fires the lamp is depicted in figure 2.

It is necessary to compensate for one-half of the delay of the output voltage behind the input voltage by introducing a mechanical lag in the connection between the comparator carriage and the wavelength dial. This is done automatically in the proper direction by means of a mechanical clutch which links the comparator drive to the wavelength drive. The wavelength dial which is to be photographed can be seen at the extreme right front of figure 1, while a similar dial provided for the convenience of the operator is seen slightly to the left of this.

The " maximum-picker ", as the electrical line-measuring device is called, does not operate on the slope of the density curve of the line, as might be supposed, for the graininess of the plate makes this impossible. Also it is made undesirable by the saturation of the photographic emulsion at high densities, which causes a flattening of the tops of dense lines. Instead the slit is set at the desired width and the line peak is taken as being half-way between two areas of equal density, measured at the points of approximately maximum slope for the narrowest line which can be resolved by the spectrograph on which the plate was exposed.

The wavelength recorder consists of a 5-bank counter, connected to a large dial which gives three more figures. Figure 3 shows the appearance of a portion of a wavelength record. The large dial is marked with a hundred divisions, and is numbered every five divisions to insure that at least one set of figures

will always appear in the field of view. This dial can be run at 1200 revolutions per minute and, as ordinarily used, one turn of the dial corresponds to 0·1 A.

The wavelength reading is recorded with a single flash of a U-shaped glass lamp containing argon at several mm. mercury pressure. A condenser is discharged through the lamp when the operator presses a key, or, in automatic recording, when the control voltage from the " maximum-picker " falls to zero after having had a positive value.

Though the relation between distance along the plate and the wavelengths

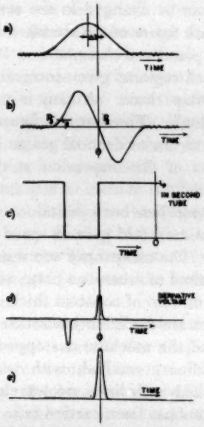


Figure 2. Curve showing the variation with time of the currents in various parts of the " maximum-picker " circuit. (*a*) The density contour of the line in terms of current from the electron multiplier. (*b*) The voltage difference between the input and output of the delay network. The line " peak " is at 0, but actually there is a delay Δt which must be compensated mechanically. (*c*) The variation in plate current i_p at the second tube of an amplifier operated by a thyratron to which is applied the voltage of (*b*). This current is sent through a transformer giving the negative and positive pulses as shown in (*d*). The positive pulse only is amplified, as in (*e*), and operates the stroboscope lamp.

of the lines recorded thereon is not linear, the instrument has been arranged to record wavelengths directly in angstroms. The process of conversion from one to the other consists merely of the four basic arithmetical operations, and these can all be performed automatically by means of differentials to provide addition and subtraction, integrators (variable speed units) to provide multiplication and division, and a few gears.

Seven-figure precision is made possible by the fact that the dispersion produced by large diffraction gratings varies not more than 2% over a 20-inch region. By

giving to different sets of gears control over the various digits representing the dispersion the machine can be provided with a dial which always reads the wavelength corresponding to the position on the plate which is passing across the centre of the slit. A cam to vary the dispersion appropriately is laid out initially by measuring standard lines. Then plate after plate, taken in the same region of the same spectrograph, can be measured with this cam. Small variations will be found when plates are changed, but these can be reduced to any desired value. Variations as great as 0·010 A. may be conveniently left in, since the standard lines are recorded on the film with the unknowns. Thus wavelength values read from the film can be changed in the seventh figure to accord with the error found in the seventh figure of the standard line readings.

The track to which the plates are clamped in a 10-metre Paschen mounting is divided into fifteen 20-inch regions, given integral numbers, and fifteen half-integral regions which overlap these. A cam is provided on the measuring engine for each of these regions. The first two figures of any needed dispersion are set into the machine by means of decimal gears. Then the appropriate cam keeps the remaining figures of the dispersion at the instantaneously desired value.

The automatic comparator has been found to possess several advantages beyond the more than a hundred-fold gain in speed of measurement and computation which it provides. The comparator screw does not change temperature during the two minutes required to measure a plate, and its nut is pushed forward at uniform speed so that an oil film of constant thickness is kept in the screw-nut contact. Wavelength values are available instantaneously, so that any line can be identified by inspection if the machine is stopped or is operated by hand—a convenience over the ordinary method, with which, after several days of measuring a spectrogram filled with lines, no wavelengths are obtained until a lengthy series of computations has been carried out.

It is found that the reproducibility of photo-electric setting on narrow lines is from three to ten times more precise than that of eye setting. Spectrum lines of various breadths were picked by the machine thirty times in each direction, and then the same lines were set on by an experienced operator. The average internal deviation of the hand-and-eye readings was $\pm 0{\cdot}002$ A., while that of those taken by machine was $\pm 0{\cdot}0004$ A.

Two automatic comparators, each capable of measuring plates up to 24 inches in length, are now in operation. More than twenty million measurements have been made with these machines, and as a result of their having been under continual improvement over several years, most of their weak points have been eliminated. They omit readings for lines which lie on the shoulders of very strong lines in such a way that no real minimum of density exists between the two lines, but in such cases it is possible to determine wavelength values with a ruler from the density trace which is incorporated in the record. Exhaustive tests have shown that errors of other sorts, arising from standards of wavelength

and variations in line shape, greatly outweigh errors introduced by the automatic features of the comparator. Attachments have been devised which will make possible the measurement of plates of low dispersion in which the dispersion is varying rapidly, but thus far use of the direct-reading wavelength feature has been confined to plates in which the dispersion variation is less than 2 % along the length of the plate.

§ 4. THE WAVELENGTH PROJECT

When the first automatic comparator became available in 1935, a programme of wavelength measurements was embarked upon. A project was initiated for which the Works Project Administration has furnished clerical aid over a five-year period. In 1939 some of the results of this work were published in the *M.I.T. Wavelength Tables*, which contain the intensities in arc, spark, or discharge tube of the 109,000 strongest spectrum lines in the range 2000 to 10,000 A. These Tables are limited to lines previously identified by other workers as originating from normal atoms or atoms in their first stage of ionization, lines from higher stages of ionization being omitted.

It was desired in this catalogue to include the best-known wavelength values for all lines, from whatever authority, so wavelength values were taken from the literature where these were deemed superior to those obtained from the Wavelength Project. Though the Tables contain only the strongest half of the known spectrum lines in the region covered, the lines included account for 99 % of the radiation emitted.

In preparing the spectrum catalogue of which the published Tables are one part, the literature was combed carefully for all published wavelength values. Data regarding each line were recorded on one of 280,000 white cards, which listed the element, the stage of ionization where known, the various measured wavelengths, the intensity of the line in arc and spark, and references to the authorities giving the data. Because the work was being done by clerks who had little familiarity with the field in which they were working, the literature research was duplicated by another group, and the results from the two groups were compared and corrected. Then all data were copied on 280,000 buff cards. The white cards were then arranged in order of wavelength under the individual elements, while the buff cards were shuffled together and arranged in order of wavelengths for all elements. The white catalogue is used to determine what lines are to be expected from a given atom or ion, while the buff catalogue is used to identify the atom or ion of origin of a given spectrum line.

When the wavelength-measuring programme was embarked upon, it soon became evident that the equipment available would be most useful in photographing the spectra of complex metallic atoms. The high dispersions used, ordinarily 0·4 A./mm., were not needed for the atomic spectra of gases, so most of our effort was concentrated on the rare earths and the elements in the long periods of the periodic table.

When twenty or thirty sets of spectrograms had been accumulated, these were grouped as to spectrum region covered, and were sent through the automatic comparators. The resulting films were then read by the W.P.A. clerks, who copied off the wavelength values, correcting the last figure in accordance with the readings of the standard lines.

After at least seven determinations from different plates had been made for each line (which resulted in more than forty measurements being accumulated for some of the lines), the wavelength values were averaged, and the average values were recorded on the white and buff cards as M.I.T. values. Whenever our values agreed acceptably with the mean of those given in the literature the cards were placed in the final catalogue. Whenever our values did not agree, new spectrograms were taken, and from seven to twenty new measurements were made on the uncertain lines. Usually the disagreement was then resolved, but in cases where it was not, a third set of measurements was made, from new plates when possible. In some cases it was found that three entirely independent sets of our measurements would agree to within $\pm 0{\cdot}003$ A., while departing by as much as $0{\cdot}04$ A. from a value given to seven figures in the literature. In such cases our own value was used in the final wavelength catalogue when we felt sure that we were measuring the proper line. When we could not be sure we used the literature value and held our own for further study.

To prepare the published *M.I.T. Wavelength Tables*, critically selected data for the 110,000 strongest lines were copied on green cards, and these were then arranged in order of diminishing wavelengths.

Since the *M.I.T. Wavelength Tables* contain only lines listed in the literature (except a few thousand at short wavelengths marked *a* where the atom of origin seemed quite certain) the Tables must be considered as incomplete, but as representing the existing state of the art. The number of lines assigned to each element is still largely determined by the amount of effort that has been expended on the study of its spectra. Many thousands of lines measured on our plates were not included in the Tables because of the inadvisability of assigning a parent atom to the line until the assignment could be verified by analysis of the spectrum. Thousands of unlisted lines were found in almost every element having a complex spectrum, though many of them, to be sure, had low intensities. Even strong lines of this sort were not included in the Tables, however, because of the desirability of taking further spectrograms of arcs in various atmospheres to avoid the danger of mistaking band lines for atomic lines. However, in the range 2400 to 2000 A. most spectra had been photographed so incompletely that inclusion of a number of lines in this region, not previously listed in the literature, was deemed justifiable.

The standard Pfund arc was used for arc excitation, light ordinarily being taken from regions not closer than 7 mm. to an electrode, to avoid pole effects. This precaution was taken only at wavelengths below 4500 A., for at longer wavelengths it was found necessary to use a shorter arc to bring out the desired

lines, and pole effects are less likely to occur there. Since the wavelength of a line excited in a spark is much more variable than its wavelength in an arc, spark wavelengths were recorded only in cases where no arc wavelengths were available.

It is believed that wavelengths given to three figures after the decimal in the Wavelength Tables are correct to within ±0·005 A., those given to two figures to within ±0·05 A., and those given to one figure to within ±0·5 A. Actually those tolerances are only approximate, for the precision obtainable varies greatly with the natural breadth of the line, its simple or complex structure, its variability under different conditions of excitation, the region of the spectrum in which it lies, and numerous factors which are difficult to evaluate. In an element such as cerium, for example, the majority of the measurements seldom deviated by more than 0·003 A. from the mean, while the final average values were consistent with each other to within 0·0016 A., as shown by application of the combination principle.

The intensities of the lines were estimated from the original plates, and from the density trace. A scale of 1 to 10,000 was used for intensity estimates, lines under 2 being omitted from the Tables.

During the first year of the project all plates were measured by hand setting as well as by automatic setting. When setting by hand the operator observed the spectrum projected on a screen before him, and when he had set the desired line on the fiducial mark he pressed a key which recorded the wavelength. This process, though considerably slower than automatic recording, is still twenty times as rapid as that commonly used, since it eliminates time consumed in writing down comparator readings and in computing wavelength values from these. It was soon found that measurements made with the automatic "maximum-picker" were more self-consistent than hand settings, and ultimately complete reliance was placed on the automatic method.

It is believed that the resulting tables are superior in precision and inclusiveness to any other tables designed for the identification of spectrum lines, yet they still suffer from serious limitations. The wavelength precision of the measurements was limited by the following factors, arranged in what is believed to be the order of decreasing importance:—

1. Variations of wavelength with excitation conditions.
2. Lack of satisfactory wavelength standards in some spectral regions, particularly of standard lines of fairly uniform intensity.
3. The displacement on spectrograms of lines to be measured, relative to standard lines.
4. Displacement caused by strong neighbouring lines, impurity lines, or bands.
5. The natural breadths of some lines and the complex structures of others.
6. Errors of coincidence inadequately corrected for where lines to be measured in one order were compared with standards in another order of the grating.

7. Limited amounts of certain materials available for excitation to produce the lines.

8. Incorrect identification of lines with lines given in the literature

We have found considerable difficulty in obtaining smooth correction curves at the high dispersions used in this work. This is because many of the international secondary standards of iron are lines of great intensity, separated in the spectrum more widely than is desirable, and entirely absent in certain regions of the spectrum. At wavelengths shorter than 2500 A. and those longer than 7000 A. interferometer measurements of iron lines made by Meggers and Humphreys* were used as standards, while the copper standards of Shenstone (1936) were used at short wavelengths.

The chief limitation of the Tables, however, arises from the fact that energy-level classifications for many of the atoms are still lacking, and thus it was impossible to determine identifications with certainty. For many of these classifications better wavelength data were needed. Therefore, it became obvious that the job should be done in two steps. The wavelength project has provided a vast mass of new material which makes possible classification of a number of spectra which have previously resisted attack. When these spectra have been classified, new wavelength standards can be computed, which should then justify a new attack on the wavelength problem and publication of a far more complete set of tables. Future tables should also contain the wavelengths of lines in higher stages of ionization than the first, and complete tables, as contrasted with those designed for analytical purposes, should of course contain the fainter lines as well as the lines contained in the present catalogue.

It is believed that the great store of plates now available can be re-worked in terms of new standards when classifications are more complete. The cerium spectrum and the thorium spectrum are two in which very precise wavelength measurements can be made with gratings. Their lines are of much more uniform intensity than are those of iron, and are spread more uniformly throughout the spectrum. The entire photographic range in air could probably be covered with standards from these two elements.

§ 5. CLASSIFICATION OF SPECTRA

Determinations of the energy levels of atoms from their line spectra are ordinarily carried out with the use of three powerful tools: the Ritz combination principle, by which every spectrum line is identified as originating from an upper and a lower energy level whose wavenumbers can be determined; Zeeman-effect studies, which reveal almost uniquely the quantum numbers of the levels from which a line originates; and excitation effects, by which lines can be grouped roughly in accordance with the energy values of their upper states. Hyperfine structure studies can also occasionally be used, though these are likely to be useful only for verification.

* Communicated in manuscript.

A number of spectrum analyses have been made in which the combination principle was the only guide. This is likely to be an uncertain process in a complex spectrum, however, owing to the number of accidental coincidences which may occur when the uncertainty of the wavelengths is great relative to the number of lines. Several classifications given in the literature are entirely spurious, and it would be difficult to over-emphasize that precise wavelength values are of the highest importance for classification of a complex spectrum.

Application of the combination principle to analysis is usually a long, slow, and tedious process. All of the lines believed to arise from a given stage of ionization of an atom are listed in accordance with their wavenumbers, and the

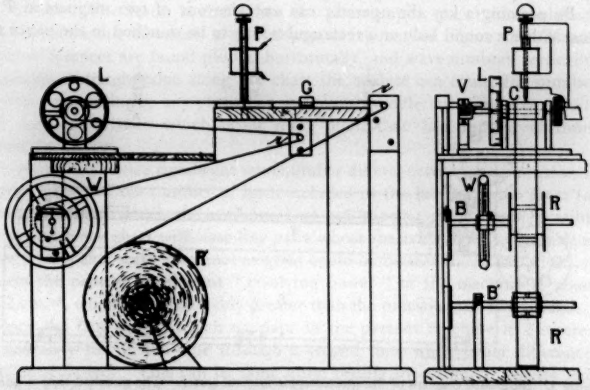


Figure 4. The punch for recording wavenumbers to seven figures on 35-mm. paper tape. The operator turns a handle at W, and the linimeter L measures distances along the tape, recording these in terms of wavenumbers. The electromagnetic punch P can be operated to punch a hole in the tape at a desired wavenumber.

wavenumber of each line is subtracted from that of each neighbouring line. This process is repeated for the various lines until wavenumber differences have been found which occur more often than would be expected by chance. All lines which show pairs of constant differences are then fitted into a quadratic array, which is gradually extended until most of the important lines can be assigned to rows and columns in the array. These rows and columns then represent levels in the atom.

To carry out these tedious seven-figure subtractions automatically and rapidly, the interval sorter was designed. The apparatus consists of two parts. The wavenumber recorded is illustrated in figure 4, and the interval sorter

itself in figure 5. A 500-ft. roll of opaque paper tape is provided, 35 mm. in width and perforated along both edges as in figure 6*a*. This roll of tape is placed on the spindle at R (figure 4) and then passes over a " linimeter " at L, which is an accurately turned aluminium cylinder (in the latest model, a sprocket wheel) running on ball-bearings and rigidly connected with an aluminium wheel on the edge of which are engraved one hundred equal divisions. A counter records the number of revolutions of the " linimeter " shaft. By suitably adjusting the diameter of the cylinder C, any desired scale of wave-numbers can be used, so that the last figure of the dial will read tenths or hundredths of reciprocal centimetres. Ordinarily one centimetre along the tape corresponds to one wavenumber.

By pressing a key the operator can energize one of two magnets at P and cause either a round hole or a rectangular hole to be punched in the paper tape,

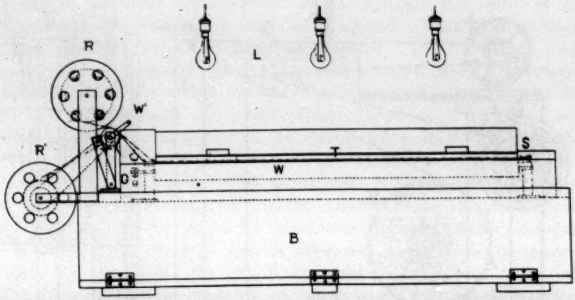


Figure 5. Simplified diagram of the spectroscopic interval sorter. The punched tape is unrolled from reel R and passes along the top of the slotted box at W for a distance of 1 metre. There it is looped over the roll at S and returns under itself to be rolled up on reel R'. Box B contains photographic paper which is moved at right angles to the motion of the tape W, recording the wavenumber differences.

as in figure 6*a*. The operator can punch several hundred holes an hour, corresponding in position to the wavenumbers of the lines in the spectrum list. The holes representing spectrum lines are rectangular, while those which are round represent fiducial marks to give a scale of wavenumber differences.

The punched tape, when completed, is transferred from the reel R of figure 4 to the reel R of figure 5. The free end of the tape is threaded through the necessary guides and rollers and passed the length of the machine, which is slightly more than a metre. The tape then passes round toothed rollers at S and returns under itself, and directly over a sheet of photographic paper which is one metre wide. After passing under another toothed roller the tape is coiled on reel R.

Above the overlapping sections of the tape are suspended electric lamps, light from which shines down through the tapes on the photographic paper wherever and whenever a hole in the upper tape passes across a hole in the lower. The long paper tape is run rapidly through the machine, about three minutes being required for 500 feet. The roll of photographic paper, one metre wide, moves at right angles to the motion of the tape, at any desired constant speed.

Since distance along the tape represents wavenumbers, when a hole in the upper tape coincides with a hole in the lower tape the position of the resulting dot on the photographic paper measures the distance along the loop of tape between the two holes, and hence records their difference in wavenumber. The motion of the photographic paper at right angles is uniform in terms of wavenumber. The result is that when the photographic paper is developed, a chart is obtained, of which a typical part is shown in figure 6*b*. Wavenumber differences are found plotted horizontally, and wavenumbers vertically. By running a straight-edge along the chart the analyst can count the number of dots corresponding to any particular wavenumber difference, and can readily locate those differences which occur more often than the number of times expected by chance.

By use of this device significant wavenumber differences become immediately apparent, provided the number of lines included in the list is not too great for the precision with which the wavenumbers are known. It is found possible to detect from the chart with ease line pairs whose intervals differ by as little as 0·1 cm.$^{-1}$, and as a scale five times as great could be used without complication, this sets the present convenient " resolving power " of the machine at about $\pm$ 0·02 cm.$^{-1}$, which is considerably greater than the precision usually needed.

Since the tape length which overlaps in the present machine is 2 metres, it is necessary to run the tape through a second time when larger differences are to be recorded. This can be done quite readily by setting a 2-metre loop at the back of the machine and running the tape through again. Loops up to 10 metres in length have been used successfully.

The function of the interval sorter may be described as the rapid concentrating of the raw material of quadratic arrays. Every dot on the record represents a subtraction of two seven-figure numbers, most of which subtractions are not needed. This saves the amateur, who might carry out almost every possible subtraction indicated, about 99 % of his time, while the expert, who would restrict his operations to what might correspond to a band perhaps ten wavenumbers wide across the record, is saved 90% of his time. The machine narrows down this band to its ultimate useful width.

In the original model of the interval sorter, the rear sprocket which controlled the length of loop, and thus the range of intervals being recorded, was driven directly by the tape itself. Occasionally the two sections of tape would interlock with resulting breakage, but this has been eliminated by driving the rear sprocket from the front sprocket through a connecting shaft.

In order to identify definitely the two lines which combine to produce a particular dot on the chart, it was necessary to use one or two metres of photographic paper to obtain a sufficiently open frequency scale. A better way of doing this was furnished by the development of the interval recorder, shown in figures 7 and 8. A punched tape similar to that used in the interval sorter, though now containing all of the lines in the spectrum list, is fed through the machine. From an upper sprocket connected to a wavenumber dial the tape passes through a loop of from 8 inches to 30 feet in length to a lower sprocket, from which it is coiled up on a reel. A beam of light from a small lamp is

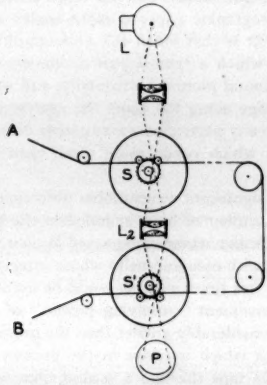


Figure 7. Diagram illustrating the principle of the spectroscopic interval recorder. Tape passes from A to B and dials S and S′ record the wavenumber readings of the two positions through which the light beam passes. Light from L passes through the upper and lower sections of the tape when holes pass the upper and lower positions simultaneously, and falls on a photocell P, which operates a stroboscope lamp giving an instantaneous photograph of the two dials.

focused on the tape, the width of this beam being adjusted to correspond to the wavenumber spread which it is desired to include under a given wavenumber difference. The diverging light beam, after passing through the upper tape, is focused by a lens on to the lower tape, and then diverges again to fill a phototube.

When holes in the upper and lower tape-sections are in line light falls on the phototube. This trips a thyratron, which discharges a condenser through an argon-filled U-tube, similar to that used in the automatic comparator, and this stroboscope lamp illuminates the wavenumber dials, which move in synchronism with the tape. A small camera records on motion-picture positive film the wavenumbers corresponding to any two holes which pass together

through the light beam. The two sprocket shafts are connected by a differential which carries a dial calibrated to read the wavenumber difference between the two dials, and this differential can be set to and clamped at any desired difference, thus controlling the length of the loop.

Figure 9 shows two short sections of records obtained with the interval recorder, which has been used to date in the analyses of more than a dozen spectra. Once the tape has been punched, a few minutes suffice to comb a list of several thousand lines for a given interval, and one tape will serve for several hundred intervals before it wears out. As many as 150 line pairs showing the same interval within a narrow tolerance may occur in a spectrum as complex as that of Ce II, though some of these are, of course, the result of chance.

The results obtained with the interval sorter and the interval recorder are combined to set up a quadratic array, thus eliminating much of the drudgery from application of the combination principle to the analysis of spectra.

§ 6. ZEEMAN-EFFECT STUDIES

It appears to be true that the combination principle alone is not sufficiently powerful to make readily possible the analysis of spectra so complex as those of some of the rare earths, even when the most precise wavelength measurements obtainable with diffraction gratings are used. Hundreds of quadratic arrays can be set up, but to tell which of these is the correct one it is necessary to make use of other data such as excitation potentials, hyperfine structure, or, best of all, Zeeman effects. The clearly resolved Zeeman pattern of a line reveals almost uniquely the quantum numbers of the two states from which it arises. Though the study of Zeeman effects affords one of the most powerful aids to the classification of spectra, it has been limited in great degree by the relatively weak magnetic fields which could be maintained over appreciable periods, by the small volume and lack of uniformity of these fields, and by the low intensities of light sources operated in the fields of iron-core electromagnets. Most of the Zeeman studies reported in the literature have involved fields of 43,000 oersteds or less, obtainable with commercial iron-core magnets of the Weiss type.

We have used the Bitter magnet successfully to resolve Zeeman patterns of complex spectra at fields up to 99,830 oersteds. This magnet will accommodate a bright electric arc (Harrison and Bitter, 1940), with which exposures of from five to thirty minutes suffice to give dense spectrograms, even with gratings of resolving power 100,000 and over. This magnet and some patterns obtained with it have been described and illustrated in a previous volume of this series (White, 1940), and needs only a brief discussion here. It consists of a coil composed of 200 turns of copper strip through which currents up to 10,000 amperes can be sent continuously, the coil being enclosed in a bronze jacket. Heat is removed by water which flows past the coil at the rate of 800 gallons per minute. A 1700-kilowatt motor generator set furnishes d.c. power at voltages up to 170.

The light source is placed within a central tube 3 cm. in diameter, in the centre of which the magnetic field is constant to 1 % over a range of 4 cm. The current through the magnet coils is kept sufficiently constant by manual control, the field varying by not more than $\pm 0{\cdot}1$ % for hours at a time.

One of the greatest advantages of this type of electromagnet for Zeeman-effect work is that the ion stream in the electric arc moves parallel to the lines of force, so that much of the quenching action of the field is eliminated. An increase in excitation results, however, when the arc is burned in a magnetic field. Unpublished results obtained by Corliss in this laboratory indicate that it may be possible to offset this increase in excitation by increasing the pressure of the gas in which the arc burns, though this has not yet been tried with a large magnet.

Spectrograms covering the range 2500 to 8000 A. have been taken for the following elements: cerium, chromium, cobalt, columbium, dysprosium, europium, gadolinium, iron, lanthanum, manganese, molybdenum, neodymium, nickel, osmium, palladium, praseodymium, rhenium, rhodium, ruthenium, samarium, tantalum, thorium, thulium, titanium, tungsten, uranium, vanadium, ytterbium, yttrium, zirconium.

Data from these plates are being accumulated as rapidly as possible. Practically all have been measured with the automatic comparators, and their films are in storage with the plates. Many of the films have been read by W. P. A. clerks, but the work has proceeded more slowly in recent months and is now being suspended for the duration of the present emergency.

Some of the plates and films have been turned over to other workers in the field, notably those for: europium, to H. N. Russell, W. E. Albertson and D. N. Davis (1941); tungsten, to J. H. Roberson, J. E. Mack and G. R. Harrison (1940); chromium and molybdenum, to C. C. Kiess at the National Bureau of Standards; iron, to D. W. Weeks at Wilson College; columbium and thulium, to W. F. Meggers at the National Bureau of Standards. Other data have been published in a series of papers from this laboratory on ruthenium (Harrison and McNally, 1940), rhodium (Molnar and Hitchcock, 1940), cerium (Harrison, Albertson and Hosford, 1941), praseodymium (Rosen, Harrison and McNally, 1941), neodymium (Albertson, Harrison and McNally, 1942) and palladium (Lindsley and Rosen, 1941), and it is expected that data for nickel, lanthanum, titanium, thorium and tantalum will be published before long.

In complex spectra, where several lines may be found per angstrom, the Zeeman patterns often overlap to such an extent as to make many of them difficult to reduce. For some elements we have taken as many as eight 12-plate sets of spectrograms at different fields, but still there are many patterns whose interpretation is uncertain. Nevertheless, the g-values obtained from the patterns which have been reduced in most cases show remarkable consistency even in the third figure after the decimal place. Russell, Albertson and Davis

(1941) have calculated that the probable error in g from a single pattern on the Eu plates supplied to them is $\pm 0{\cdot}005$, and for the final averages from several lines $\pm 0{\cdot}0011$ is the probable error in many cases. We have found even higher precision in other elements.

In the present investigation of the films only the most obvious data are being reduced. These involve determination of the g-values of as many levels as possible, each from as many lines as possible, to aid in the classification of the spectra. A vast amount of data remains on the films for reduction when they can be studied more carefully.

§ 7. NEW SPECTRA CLASSIFIED

The availability of the new wavelength and Zeeman-effect data has made possible successful attacks on many spectra previously found impregnable. We are now very optimistic regarding the possibility of the analysis of the principal lines of any spectrum, even those of the rare earths, though, owing to the fact that most of the Zeeman patterns obtained for the rare earth elements are for the singly ionized atoms, we have not yet had much success in attacking such spectra as Ce I, Nd I, and Pr I. However, beginnings have been made on these, and it is hoped that the perfecting of an arc to run at high pressures in the magnet will make possible a successful extension of the analysis to these important spectra.

Plans for further improvements in tools to aid this work must be postponed temporarily. Among those planned and partly constructed is an automatic machine for the reduction of Fabry-Perot fringes. It has been found that modern high-intensity gratings can be crossed with Fabry-Perot quartz etalons to give strong exposures over broad regions of the spectrum in a short time. A device has been developed to change wavelengths rapidly to wavenumbers, and another device should make possible automatic reduction of patterns and a thousand-fold increase in the speed of interferometer measurements. At the moment such developments must wait.

REFERENCES

ALBERTSON, W. E., HARRISON, G. R. and MCNALLY, J. R., Jr., 1942. *Phys. Rev.* **61**, 167.
BITTER, F., 1939. *Rev. Sci. Instrum.* **10**, 373.
BITTER, F. and HARRISON, G. R., 1940. *Phys. Rev.* **57**, 15.
HARRISON, G. R., 1932. *Rev. Sci. Instrum.* **3**, 753.
HARRISON, G. R., 1933 a. *Amer. Phys. Teacher*, **1**, 109.
HARRISON, G. R., 1933 b. *Rev. Sci. Instrum.* **4**, 581.
HARRISON, G. R., 1933 c. *Rev. Sci. Instrum.* **4**, 651.
HARRISON, G. R., 1935. *J. Opt. Soc. Amer.* **25**, 169.
HARRISON, G. R., 1938 a. *Rev. Sci. Instrum.* **9**, 15.
HARRISON, G. R., 1938 b. *J. Opt. Soc. Amer.* **28**, 290.
HARRISON, G. R., ALBERTSON, W. E. and HOSFORD, N., 1941. *J. Opt. Soc. Amer.* **31**, 439.
HARRISON, G. R. and BITTER, F., 1940. *Phys. Rev.* **57**, 15.
HARRISON, G. R. and MCNALLY, J. R., Jr., 1940. *Phys. Rev.* **58**, 703.
HARRISON, G. R. and MOLNAR, J. P., 1940. *J. Opt. Soc. Amer.* **30**, 343.

LINDSLEY, H. and ROSEN, N., 1941. *J. Opt. Soc. Amer.* In press.
MOLNAR, J. P. and HITCHCOCK, W. J., 1940. *J. Opt. Soc. Amer.* **30**, 523.
ROBERSON, J. H., MACK, J. E. and HARRISON, G. R., 1940. *Phys. Rev.* **58**, 895, and other papers in preparation.
ROSEN, N., HARRISON, G. R. and MCNALLY, J. R., Jr., 1941. *Phys. Rev.* **60**, 722.
RUSSELL, H. N., ALBERTSON, W. E. and DAVIS, D. N., 1941. *Phys. Rev.* **60**, 641.
SHENSTONE, A. G., 1936. *Philos. Trans.* A, **235**, 145.
WHITE, H. E., 1940. *Rep. Prog. Phys.* **6**, 145.

MOLECULAR ELECTRONIC SPECTRA, DISPERSION AND POLARIZATION

THE THEORETICAL INTERPRETATION AND COMPUTATION OF OSCILLATOR STRENGTHS AND INTENSITIES

By ROBERT S. MULLIKEN

AND

CAROL A. RIEKE,

Ryerson Physical Laboratory, University of Chicago

§1. INTRODUCTION

THE theoretical computation of absolute transition probabilities for electronic spectra of molecules has been attempted only recently, though theoretical formulas for relative intensities of lines in a band, and methods for calculating theoretically the relative intensities of bands in a band-system, are well known. The quantum-mechanical calculations on absolute intensities are, for most molecules, necessarily approximate; their usefulness is still being explored and the methods of computation improved. Fortunately, even very rough theoretical calculations prove to be valuable, since the transition probabilities vary over a range of several powers of ten. The methods of computation, and the results so far achieved in the actual determination of intensities and their interpretation in terms of molecular structure, form the principal subject matter of this *Report*. Applications to various other molecular properties such as dispersion, polarizability, and the interpretation of molecular refractivities, and to dyes and to the abundance of molecules in astronomical sources, are also discussed.

In the discussion of transition intensities in this *Report*, we shall consider, for the most part, only absorption spectra, since it is not feasible to obtain extensive experimental data on absolute transition probabilities in emission spectra.

§2. GENERAL EXPRESSIONS FOR TRANSITION PROBABILITIES*

(a) *Dipole and oscillator strengths*

There are two quantities either of which may be used as a convenient measure of absorption strength: the " oscillator strength " f of dispersion theory, and the " dipole strength " D. Each of these quantities may be computed quantum-theoretically, or obtained in terms of experimental absorption coefficients. In

* See Mulliken, 1939 a.

this ***Report*** we shall frequently have occasion to compare computed and observed values of f and D.

The dipole strength D for a transition between two electronic states with orbital electronic wave functions ψ_k and ψ_l may be defined* by

$$D = GQ^2, \text{ with } \boldsymbol{Q} = \int \psi_k (\Sigma_i \boldsymbol{r}_i) \psi_l d\tau. \qquad \ldots\ldots (1)$$

Unless otherwise stated, we shall, in the following discussion, always use wave functions corresponding to nuclei fixed in their ***equilibrium configuration*** for the ***initial state*** ψ_l. The vector $\boldsymbol{r}_i$ of equation (1) is the radius vector of the ith electron referred to a system of co-ordinate axes fixed in the molecule. The vector $\boldsymbol{Q}$ is conveniently obtained by computing Q_x, Q_y, and Q_z and adding vectorially. If the molecule has one or more axes of symmetry, and the co-ordinate axes are taken along and perpendicular to such axes, all but one of the integrals Q_x, Q_y, Q_z vanish; the $\boldsymbol{Q}$ vector for a given pair of wave functions ψ_k, ψ_l is then directed along one of the co-ordinate axes. It is assumed here and throughout the ***Report*** that only real wave functions and real co-ordinates are being used.

In case either or both of the wave functions ψ_k and ψ_l belong to degenerate energy levels, there are two or more ψ_k's and/or ψ_l's and therefore two or more $\boldsymbol{Q}$ vectors for a given energy-level transition. It can be shown, however, that in all such cases the different $\boldsymbol{Q}$ vectors are equal in magnitude. For the total ***intensity*** of a transition of given frequency from a single initial wave function ψ_l, the effect of the existence of more than one ψ_k is therefore merely to introduce the factor G which appears in the expression for D in equation (1). This factor is equal to the number of independent ψ_k's to which transition can occur from any one ψ_l. Orbital degeneracy in the initial state l need not be considered in equation (1), since it merely causes a distribution of absorbing molecules among different ψ_l's, but does not affect the absorbing power per molecule.

It should be noted carefully that equation (1) gives the dipole strength for a transition between two ***orbital electronic*** energy levels k and l. The fact that the levels k and l are not really single, but that each consists of a complicated family of levels, resulting from molecular vibration and rotation and often also electron spin, is here ignored. This, however, should lead to no serious error provided the theoretical D from equation (1) is compared with an experimental D based on a summation of intensity over all bands, band-lines, and spin components (including intersystem transitions) which may exist in the observed spectrum associated with the orbital electronic change $l \rightarrow k$. This can be shown as follows.† For any given individual initial state (definite ψ_l, and definite spin, rotational, and vibrational wave function), D of equation (1) gives the sum total of dipole strengths of all possible transitions to the various ψ_k, spin, rotational, and vibrational states of the *final* orbital electronic energy-level k.

* In Mulliken, 1939 a, D was defined by $D = Q^2$, and a factor G' was included in equations corresponding to our equations (2) and (4). It is more convenient in practice to include G' in equation (1).

† Cf. Mulliken, 1939 a, pp. 15–17.

Further, since the total strength D is the same for every one of the different possible individual states associated with the *initial* level l, it does not matter how the molecules are distributed among these individual initial states; the total D per molecule for a mass of absorbing molecules is unaffected. An important corollary is that the experimental D should be independent of temperature.

Deviations from the sum rule for D used in the foregoing discussion are theoretically possible, and in fact necessary, but are unimportant here in view of the roughness of the theoretical computations. The principal cause of such deviations is the fact that ψ_k and ψ_l of equation (1) are not entirely independent of nuclear positions (cf. § 9 (a)). It is thus seen that our procedure of using ψ's corresponding to nuclei fixed in their initial equilibrium configuration can cause slight errors. Large errors are not usually to be expected because, according to the Franck-Condon principle, vibrational jumps, in which (in the first instant) the nuclei go far from their initial equilibrium configuration, are relatively improbable, thus contributing relatively little to the total D.

Let us assume that the electron jump in which we are interested is represented by a single narrow absorption band at wavenumber ν.* The following relation can then be shown to exist:

$$f=(8\pi^2\boldsymbol{mc}/3\boldsymbol{h})\nu D=1{\cdot}085\times10^{11}\nu D, \qquad \ldots\ldots(2)$$

where $\boldsymbol{m}$ and $\boldsymbol{c}$ are the mass of the electron and the speed of light, and $\boldsymbol{h}$ is Planck's constant†. The f's for any atom or molecule obey the sum rule $\Sigma f=Z$, where Z is the total number of electrons present.

Equations (1) and (2) give relations of D and f to quantum-mechanical wave functions. On the other hand, their relations to experimental quantities are given in terms of absorption coefficients, k_ν, defined by

$$I_\nu=I_\nu^0e^{-k_\nu l},$$

where l is the length of absorbing path in cm., and I_ν^0 and I_ν are the light intensities at wave number ν before and after absorption. For molecules, most experiments have been made in solution, and instead of k_ν it is then customary to employ the molecular extinction coefficient ϵ_ν defined by

$$I_\nu=I_\nu^0 10^{-\epsilon_\nu cl},$$

where c is the concentration in moles per litre. For a *gas* at 0° C. and one atmosphere, $c=1/22{\cdot}4$ and $k_\nu=0{\cdot}1028\,\epsilon_\nu$.

The oscillator strength is related to the absorption coefficient by

$$f=(\boldsymbol{mc}^2/\pi\boldsymbol{e}^2N)\int\gamma k_\nu d\nu=4{\cdot}20\times10^{-8}\int\gamma k_\nu d\nu, \qquad \ldots\ldots(3)$$

* Here and elsewhere throughout this *Report*, ν is in *wavenumbers*.

† The numerical values of constants are based on the values as compiled by Birge (*Phys. Rev.* **52**, 241 (1937) and earlier papers). (See also the *Report* by Birge, p. 90 of this volume—[Editor].)

where N is the number of molecules per c.c., and the factor γ is introduced to allow absorption data on liquids or solutions to be used. (For a gas, $\gamma = n = 1$, where n is the index of refraction.) Similarly,

$$D = (3hc/8\pi^3 Ne^2)\int \gamma k_\nu d\nu/\nu = (3{\cdot}88 \times 10^{-19})\int \gamma k_\nu d\nu/\nu. \quad \ldots\ldots(4)$$

With the inclusion of γ, equations (3) and (4) yield experimental f and D values referring to isolated or vapour state molecules.

Equations (2), (3), (4) are strictly applicable only if the spectrum associated with the electron jump is confined to one very narrow band, but may be used with sufficient accuracy for our purposes even when it is spread out in an extended system of bands. For this purpose, ν in equation (2) may be taken as a weighted mean ν for the spectrum, or, approximately, as ν_{max}, i.e. the ν for peak absorption. In equations (3) and (4) the integrations are taken over the whole of the spectrum belonging to the electronic transition in question.

To obtain a purely theoretical f, ν as well as D in equation (2) should be computed theoretically. This can often be done approximately, but it has been more usual to combine the theoretical D with the empirical ν_{max}. The f thus obtained may be called a "semi-theoretical" f. Either the purely theoretical or the semi-theoretical f may then be compared with the observed f as given by equation (3). Or, observed and computed D's may be compared.

(b) *Comparison of vapour and solution f values*

Almost all measurements of the absorption coefficients of molecules have been made in solution. This procedure has the advantage of suppressing the fine structure of the bands, which, if consisting of discrete lines, makes the accurate determination of the $\int k_\nu d\nu$ very difficult. But it has the disadvantage of introducing the factor γ, which cannot in general be accurately evaluated. Because of existing confusion as to the determination and use of γ, it seems appropriate to discuss the matter here in some detail.

The factor γ is due to the influence of neighbouring molecules on the dissolved molecules, in the form of definite chemical changes, and of other electrical effects such as the Stark effect or Lorentz-Lorenz forces. No general rules can be given concerning correction for chemical effects, which should be kept small by the use of inert solvents. Of the other effects, it is ordinarily tacitly assumed that the Lorentz-Lorenz forces are the most important. Using classical dispersion theory, Chako (1934) has found that the value of γ due to Lorentz-Lorenz forces is $9n_0/(n_0^2+2)^2$ for a solution in regions of transparency if the index of refraction is n_0 for the solvent.*

* Chako gives the correction in the form $\bar{f} = 9\bar{f}/(n_0^2+2)^2$, where $\bar{f}$ includes the factor n_0. This has unfortunately led to some misunderstanding in the application of the correction by Mulliken (1939 c, p. 131), Henri and Pickett (1939), and Child and Walker (1938), who all used the factor $9/(n_0^2+2)^2$, or about 1/2 for bands near 2500 A., instead of the factor of approximately 3/4 given by Chako's theoretical development.

It would be desirable to test the success of the Lorentz-Lorenz correction experimentally, by comparing the $\int k_\nu d\nu$ measured in vapour and in solution. Unfortunately, the necessary data are scarce, and are for the most part on transitions unsuitable for the comparison. For example, pure bromine vapour has a very weak continuous absorption at about 44,000 cm.$^{-1}$, but in liquid, in solution, and even in the presence of foreign gases at pressures of 20 to 60 cm. Hg, the absorption coefficients are increased many times (Bayliss and Rees, 1939). This provides an extreme example of disturbing influences, quite inexplicable by the simple Lorentz-Lorenz theory. In the stronger bromine absorption continuum with maximum near 25,000 cm.$^{-1}$, the effects are much smaller, but still much too large for the Lorentz-Lorenz theory: the intensities in solution are on the average twice the vapour intensities, while added foreign gases increase the intensity 10 to 25 %. Similar effects are observed in chlorine and iodine (Rabinowitch and Wood, 1936), for which the data are less complete. The very large effects of solvents on the halogen spectra may be explained as perturbing effects on transitions which (Mulliken, 1940 a) are forbidden or weak in the vapour state.

There are also vapour and solution measurements on the weak longest wave-length absorption region in acetone (Scheibe and Fromel, 1937), acrolein, and crotonaldehyde (McMurry, 1941 b), but the transition involved is believed to be one forbidden by the electronic selection rules (Mulliken, 1935 b; McMurry and Mulliken, 1940). Here the difference between vapour and solution f values is small; the values are 0·0004 and 0·0003 respectively for acrolein; 0·0004 and 0·0005 for crotonaldehyde; and for acetone the f value is about 10 % greater in hexane, 25 % in water, than the vapour value.

The most suitable data for the comparison are provided by measurements on cyclohexadiene (Henri and Pickett, 1939) and on cyclopentadiene (Pickett, Paddock and Sackter, 1941). For these the absorption corresponds to a strong allowed transition, and the solutions were in the inert solvent hexane. For cyclohexadiene, the $\int k_\nu d\nu$ was found to be practically equal for vapour and solution. For cyclopentadiene, the $\int k_\nu d\nu$ is about 20 % higher for the vapour than for the solution, an effect just opposite to that predicted on the basis of Lorentz-Lorenz forces. Lack of resolution in the vapour spectrum could of course lead to error in the $\int k_\nu d\nu$, but in the direction of making the measured vapour f values appear *smaller* than the true values (cf. Oldenberg and Rieke, 1938).

The classical Lorentz-Lorenz correction is thus clearly inadequate to explain the available data even in two very favourable cases. The true ratio of vapour to solution intensity should therefore be determined empirically in every case. In the absence of such information, it seems best to use $\gamma = 1$ as a value which, judging by the results on the two dienes, is probably roughly correct if we confine

ourselves to allowed transitions of molecules observed in inert solvents. The experimental f values used in the following discussion will then be for the most part solution f values treated as equal to vapour f values.

§ 3. GENERAL REMARKS ON METHODS FOR MOLECULAR COMPUTATIONS

The theoretical computations of absorption strengths are based on equation (1), and the main problem is the choice of suitable approximate wave functions for use in this equation. When practicable, ψ's are used and the integration usually presents no serious difficulties. In setting up the wave functions, the concept of a unique electron configuration, which has been very useful for atomic calculations, may be carried over to the more complicated molecular case (cf. Mulliken, 1935 a). But the inexactness of the electron configuration idea is sufficiently great for molecules to admit the possibility of two or more alternative configurations, each of which may form a basis for a fairly good approximation to the complete wave function. We must then adopt views of molecular structures which are more or less intermediate between those suggested by the alternative configurations.

Methods of assigning molecular electron configurations may be divided into two types. In one of these atomic orbitals* (AO's) are used exclusively; in the other, molecular orbitals (MO's) of some sort are used for shared electrons. The first method follows the concepts of chemistry and treats every molecule, so far as possible, as composed of definite atoms or ions. The electron configuration is then the sum of the configurations of these atoms or ions, e.g., the normal state of Li_2 may be represented as

$$1s^2 1s^2(2s\,.\,2s),\ {}^1\Sigma^+_g. \qquad \ldots\ldots(5)$$

So far as homopolar binding is concerned, this method, which grew out of the work of Heitler and London on H_2, was developed chiefly by Pauling and Slater. It has had notable success as a qualitative conceptual scheme for interpreting and explaining empirical rules of valence and in semi-quantitative, mostly semi-empirical calculations of energies of formation and other properties of molecular electronic states, particularly the normal states.

Departing from chemical concepts, the second method treats each molecule, so far as possible, as a unit. This method was introduced by Hund for diatomic molecules, and was developed by Hund, Mulliken, Hückel, Lennard-Jones, and others. In its pure form, the method goes too far: it is best to use molecular orbitals only for shared electrons and atomic orbitals for the others. As an example, the normal state of Li_2 appears as

$$1s^2 1s^2(\sigma_g 2s)^2,\ {}^1\Sigma^+_g, \qquad \ldots\ldots(6)$$

* The term "orbital" means "one-electron orbital wave-function" and corresponds to "orbit" of the Bohr theory.

where $\sigma_g 2s$ represents the bonding molecular orbital. The form of $\sigma_g 2s$ is as yet unspecified. In general, this would be obtained as a solution of a Schrödinger equation for a single electron in the average field of the other electrons and the nuclei. In practice the solution of this problem is generally too difficult, but a simple approximation suitable for computation can easily be given in the so-called " linear combination of atomic orbitals " (LCAO) form. Then the $\sigma_g 2s$ MO may be approximated by

$$\sigma\, 2s \sim a(2s_A + 2s_B), \qquad \ldots\ldots(7)$$

where $2s_A$ and $2s_B$ refer to $2s$ AO's of the two lithium atoms. Here and later in this ***Report***, we shall use term symbols such as $2s$ or $\sigma_g 2s$ as convenient abbreviations for corresponding analytical AO or MO forms.

LCAO forms are reasonable from the viewpoint of variation theory, and their easy extension to large molecules has already resulted in valuable information. More accurate MO's, however, should be capable of giving still better quantitative results. For example, the work of James and Coolidge (1933, 39) and of James (1934, 35) on H_2 and Li_2, and of Coulson (1937) on H_2, while not based on pure MO-configuration wave-functions, nevertheless indicates that these would give rather good approximations.

Computations on hydrogen will be given in some detail in §5 for both AO- and MO-configuration wave-functions.

§4. VALENCE SHELL AND RYDBERG SPECTRA

Molecular electronic transitions (excluding those in the x-ray region) may be divided, roughly, into intra-valence-shell transitions and Rydberg series transitions (cf. Mulliken, 1939 b). The latter are similar to Rydberg series in atoms, and as in atoms, the lowest members are often fairly intense. In the electronic levels of a molecular Rydberg series, generally one electron is excited from a valence shell orbital to what may be called a Rydberg orbital. The Rydberg orbital is much larger than the molecule, or at least much larger than the part of the molecule from which the electron comes. Thus the field of force in which such an electron spends most of its time is not far different from that of a point charge and the orbitals differ little in form from AO's.

Emission Rydberg series were first found for the unstable helium molecule. Absorption Rydberg series have been investigated rather extensively in recent years, especially by W. C. Price (Price and Tutte, 1940; Price and Walsh, 1940; and earlier papers). They have given much interesting information, notably on molecular ionization potentials and on charge distributions in molecules. Since, however, they commonly occur in the vacuum ultra-violet, where quantitative intensity measurements are very difficult and few, and since the theoretical intensity computations for them may prove difficult and have not yet been attempted, we shall not consider them further here.

Valence-shell transitions, that is, transitions between two valence-shell orbitals, occur for the most part at lower frequencies than do Rydberg transitions,

and comprise a large part of the visible and ordinary ultra-violet band spectra. They are generally of lower intensity than the first Rydberg transitions and correspond to transitions that would be forbidden in atoms, that is to say, if the molecule could be gradually separated into atoms, a valence-shell transition would gradually get weaker and finally disappear when the separation was complete.

The name V, N has been given to a particular class of valence-shell transitions (see § 5 (a)) which generally have extremely high intensities and occur at relatively high frequencies, often approximating those of the lowest Rydberg series members.† In some cases the state V may even be classified more or less equally well with the Rydberg or with the valence-shell states. There is at least one V state and one V, N transition for every molecule; the number of such states and transitions increases rapidly with the number of valence bonds.

The name Q, N has been given to a second important class of valence-shell transitions. These involve the transfer of an electron between a usually nearly non-bonding orbital of π or related type and a bonding or antibonding orbital of σ or related type. The Q, N class may best be described by a few examples. For HCl, the electron configurations of N and Q (in MO terminology, and omitting inner electrons) are (Mulliken, 1940 c)

$$N: \sigma^2\pi_{Cl}{}^4, {}^1\Sigma^+. \qquad \ldots\ldots(8)$$

$$^3Q, {}^1Q: \sigma^2\pi_{Cl}{}^3\sigma^*, {}^3\Pi, {}^1\Pi,$$

where, in LCAO approximation,

$$\left.\begin{aligned}\sigma &\sim \alpha(3p\sigma_{Cl}) + \beta(1s_H),\\ \sigma^* &\sim \alpha^*(1s_H) - \beta^*(3p\sigma_{Cl}).\end{aligned}\right\} \qquad \ldots\ldots(9)$$

For OH (λ 3064 band) (Mulliken, 1940 c),

$$\left.\begin{aligned}N&: \sigma^2\pi_O{}^3, {}^2\Pi,\\ Q&: \sigma\pi_O{}^4, {}^2\Sigma^+,\end{aligned}\right\} \qquad \ldots\ldots(10)$$

where

$$\sigma \sim a(2p\sigma_O) + b(1s_H). \qquad \ldots\ldots(11)$$

For Cl_2 (Mulliken, 1940 b),

$$\left.\begin{aligned}N&: \sigma_g{}^2\pi_u{}^4\pi_g{}^4, {}^1\Sigma^+{}_g,\\ {}^3Q, {}^1Q&: \sigma_g{}^2\pi_u{}^4\pi_g{}^3\sigma_u, {}^3\Pi_u, {}^1\Pi_u,\end{aligned}\right\} \qquad \ldots\ldots(12)$$

where

$$\left.\begin{aligned}\pi_g &\sim a(3p\pi_A - 3p\pi_B),\\ \sigma_u &\sim b(3p\sigma_A + 3p\sigma_B).\end{aligned}\right\} \qquad \ldots\ldots(13)$$

Q, N transitions are in general much less intense than V, N transitions; in the examples cited, this is a consequence of the fact that the σ_{Cl} or σ_O part of the σ orbital is predominantly $p\sigma$, and the transition $np\sigma \rightarrow np\pi$ in a single atom is forbidden. Nevertheless they are of great interest because of the relatively long wave-lengths at which they occur.

† The letter N stands for the normal electronic state, and V for a particular type of excited state.

For diatomic molecules, we may add to our description of V, N and Q, N transitions the statement that they are respectively polarized parallel and perpendicular to the figure axis.

The V, N and Q, N transitions by no means comprise all possible valence-shell transitions. Examples of some other types are as follows. In Li_2, the transition $^1\Pi_u$, $^1\Sigma^+_g$ between the states

$$N, (\sigma_g 2s)^2, {}^1\Sigma^+_g \quad \text{and} \quad (\sigma_g 2s)(\pi_u 2p), {}^1\Pi_u \qquad \ldots\ldots(14)$$

should have high intensity, because the transition is like $ns \rightarrow np$ in a single atom, as for the sodium D lines. Many examples similar to this, or intermediate between this and the Q, N type, are known. Another type is represented by the transition in N_2^+ between the states

$$N, (\sigma_u 2s)^2(\sigma_g 2p), {}^2\Sigma^+_g \quad \text{and} \quad (\sigma_u 2s)(\sigma_g 2p)^2, {}^2\Sigma^+_u. \qquad \ldots\ldots(15)$$

Types intermediate between this and V, N occur. This type is expected to have a high intensity for the same reason as in the case of the lithium transition, but in both these cases (as well as for the Q, N transition) matters are really not quite so simple as we have indicated—the MO's are actually not pure $\sigma_g 2s$, $\sigma_u 2s$, $\sigma_g 2p$, but correspond to more or less mixed forms intermediate between $\sigma_g 2s$ and $\sigma_g 2p$, or $\sigma_u 2s$ and $\sigma_u 2p$.

Certain transitions related to the V, N transitions but, in the AO though not in the MO approximation, involving simultaneous excitation of two or more electrons, may be predicted to occur with intensities much greater than for multiple excitation processes in atoms (Mulliken, 1939 b). Such transitions would, however, occur only at very short wave-lengths and have not yet been identified.

In polyatomic molecules many transitions forbidden by electronic selection rules actually occur with intensities greater than some of the allowed transitions. This type of transition will be discussed in § 9.

§ 5. INTENSITIES IN THE HYDROGEN MOLECULE SPECTRUM: DISPERSION AND POLARIZABILITY

(a) *AO and LCAO MO computations for the V, N transitions*

As an example of the AO and MO methods of computation, and in order to explain more clearly the nature of V, N transitions, we shall describe in this section in some detail the computation of the intensity for the V, N transition of the hydrogen molecule (Mulliken, 1939 b).

The electron configurations are as follows:

$$\begin{array}{ll} \text{AO approximation} & \begin{cases} \text{State } N\text{: } (1s_A)(1s_B), {}^1\Sigma^+_g, \\ \text{State } V\text{: } (1s_A)^2 \text{ or } (1s_B)^2, {}^1\Sigma^+_u; \end{cases} \\ \text{MO approximation} & \begin{cases} \text{State } N\text{: } (\sigma_g)^2, {}^1\Sigma^+_g, \\ \text{State } V\text{: } (\sigma_g)\,(\sigma_u), {}^1\Sigma^+_u. \end{cases} \end{array} \qquad \ldots\ldots(16)$$

State N is of the species $^1\Sigma^+_g$, rather similar to ^1S of an atom. State V is the lowest excited singlet state of hydrogen, of the species $^1\Sigma^+_u$, and the transition V, N is a well-known intense transition giving a system of bands near 1000 A. (the Lyman bands), observed both in absorption and in emission.

In terms of AO's, the molecule is thought of as consisting in state N of a pair of neutral *atoms* A and B, each with one $1s$ electron; this is expressed by writing $1s_A$ and $1s_B$. State V is believed to have an ionic type of structure, either H^+H^- or H^-H^+, expressed by assigning both electrons to the $1s$ AO of the same atom, either A or B (actually half the time on each, according to quantum mechanics). Thus from the AO viewpoint, the outstanding characteristic of the V, N transition is that *an electron is transferred from one atom to the other.* Hence the name " charge-transfer spectra " has been proposed for such transitions (Mulliken, 1939 b).

In terms of MO's, both electrons in state N are assigned to an MO of the species σ_g, here for brevity designated simply σ_g; this MO is rather analogous to $1s$ of an atom. In state V, one electron is still assigned to the σ_g MO, but the second has been excited to the lowest-energy MO of the species σ_u, here designated simply σ_u. This MO rather closely resembles $2p\sigma$ of a hydrogen atom, and may be considered as the first of a Rydberg series of σ_u MO's closely resembling $np\sigma$ AO's. For the present purpose, however, it will be better to consider both σ_u and σ_g as valence-shell MO's and to regard the V, N transition as a valence-shell transition, since in many molecules other than hydrogen, σ_u or its analogue do not so closely resemble Rydberg MO's.

The orbitals σ_g and σ_u are corresponding bonding and anti-bonding MO's, so that from the MO viewpoint the V, N transition is essentially the *transition of an electron from a bonding to the corresponding anti-bonding MO.* In the LCAO approximation, σ_g and σ_u are

$$\left.\begin{aligned}\sigma_g &\sim (1s_A + 1s_B)/\sqrt{2(1+S)},\\ \sigma_u &\sim (1s_A - 1s_B)/\sqrt{2(1-S)},\end{aligned}\right\} \qquad \ldots\ldots(17)$$

where $S = \int 1s_A 1s_B dv$ is the overlap integral for the AO's $1s_A$ and $1s_B$.

Approximate wave functions corresponding to the AO and MO electron configurations (equation (16)) are as follows (the arguments 1 and 2 denote the two electrons):

$$\text{AO:}\ \left\{\begin{aligned}\Psi_N &= [1s_A(1)1s_B(2) + 1s_B(1)1s_A(2)]/\sqrt{2(1+S^2)},\\ \Psi_V &= [1s_A(1)1s_A(2) - 1s_B(1)1s_B(2)]/\sqrt{2(1-S^2)}.\end{aligned}\right\} \qquad \ldots\ldots(18)$$

$$\text{MO:}\ \left\{\begin{aligned}\Psi_N &= \sigma_g(1)\sigma_g(2),\\ \Psi_V &= [\sigma_g(1)\sigma_u(2) + \sigma_u(1)\sigma_g(2)]/\sqrt{2}.\end{aligned}\right\} \qquad \ldots\ldots(19)$$

A factor representing the spin wave function has been omitted in equations (18) and (19), since it disappears in the following integrations.

We are now ready to use equation (1) to compute the dipole strength of the V, N transition. Taking the symmetry axis of the molecule as z axis, and noting

that $G=1$ and that the $\mathbf{Q}$ vector here reduces to Q_z (V, N being a parallel-type transition with non-degenerate upper level), we have

$$Q_{V,N}=\int \Psi_V(z_1+z_2)\Psi_N dv_1 dv_2. \qquad \ldots\ldots(20)$$

Introducing the MO wave functions (equation (19)) and noting that σ_g and σ_u are mutually orthogonal and may be taken as normalized, one readily obtains

$$Q_{V,N}(\mathrm{MO})=\sqrt{2}\int \sigma_g z \sigma_u dv, \qquad \ldots\ldots(21)$$

where the integration is now over the space of a single electron. Substitution of the LCAO forms for σ_g and σ_u leads to

$$Q_{V,N}(\mathrm{LCAO\ MO})=\left[\int (1s_A)z(1s_A)dv-\int (1s_B)z(1s_B)dv\right]\Big/\sqrt{2(1-S^2)}. \qquad \ldots.(22)$$

Now $\int (1s_A)z(1s_A)dv$ is simply the average value of z for the function $1s_A$, namely z_A. Since the internuclear distance r is equal to z_A-z_B, it follows that

$$Q_{V,N}(\mathrm{LCAO\ MO})=(z_A-z_B)/\sqrt{2(1-S^2)}=r/\sqrt{2(1-S^2)}. \qquad \ldots\ldots(23)$$

Similarly, using the AO wave functions of equation (18) we obtain

$$Q_{V,N}(\mathrm{AO})=Sr/\sqrt{1-S^4}. \qquad \ldots\ldots(24)$$

For the equilibrium value of r(0·74A.) in state N, S has the value 0·75. As a result, the two approximations happen to agree very well—Q comes out 0·80A. by equation (23) and 0·68A. by equation (24). Taking $D=Q^2$, and using equation (1) with ν taken equal to the approximate observed value 99,000 cm^{-1} of $\nu_{\max}$, semi-theoretical f values of 0·68 or 0·49 are obtained for the LCAO and AO approximations respectively. That this V, N transition corresponds to a very large oscillator strength may be appreciated by comparing the present computed f values with the value $f=0{\cdot}98$ for the sum of the two very strong sodium D lines.

(b) *Further computations on Lyman and Werner band intensities, and comparison with experiment*

It is of interest to compare with experiment the f values computed above for the Lyman bands. No quantitative absorption coefficient measurements on these bands exist, but dispersion measurements may be used instead to obtain experimental values.

In general the index of refraction n of a gas in regions of transparency is given by the theoretical expression

$$n-1=(e^2N/2\pi mc^2)\Sigma_j f_j/(\nu_j^2-\nu^2), \qquad \ldots\ldots(25)$$

where N is the number of molecules per cc. Empirically, the available refraction data for hydrogen can be represented very closely (see Wolf and Herzfeld, 1928) by

$$n-1=1{\cdot}21\times 10^{8}\{0{\cdot}84/[(106{,}100)^2-\nu^2]+0{\cdot}69/[(136{,}100)^2-\nu^2]\}. \qquad \ldots\ldots(26)$$

The representation of the dispersion data as a sum of only two terms means, of course, that several neighbouring transitions have been lumped together with an f value approximately equal to their sum, and with an average effective ν. The Lyman bands have ν_{max} at about 99,000 cm^{-1}, the Werner bands at about 104,000 cm^{-1}, while numerous weaker band systems occur at somewhat higher frequencies. The term in equation (26) with $f=0{\cdot}84$ should then correspond to the sum of all these, with the strong Lyman and Werner bands predominant. Similarly, the second term in equation (26) must correspond to the sum of further terms in equation (25) with frequencies mainly above ionization. The fact that the sum of the two f's in equation (26) falls below the value 2 required for the sum of the true f's can readily be understood as a consequence of the lumping of terms extending over a range of frequencies in equation (25) into the two terms of equation (26).

Table 1. Computed D and f values for electronic transitions in hydrogen

	Approximation used			Results[d]		
Upper state	$Z(1s)$[a]	MO configuration[b]	AO configuration[c]	D in A^2	$\nu\times10^{-3}$	f
$V\,^1\Sigma^+_u$	1		$1s^2$ (ionic)	0·46	99	0·49[e]
	1	$\sigma_g1s\sigma_u1s$		0·63	99	0·68
(Lyman bands)	1	$\sigma_g1s\sigma_u2p$		0·41	99	0·44
	1·2	$\sigma_g1s\sigma_u2p$		0·22	99	0·24
	1		$1s2p\sigma$	0·36	99	0·39
	1		$1s2p\sigma$	(0·17)	(105)	(0·20)
	1·2		$1s2p\sigma$	0·19	99	0·21
	1		$1s2s$	0·043	high	
	1		$1s2s$	(0·031)	(102)	(0·035)
$^1\Pi_u$	1	$\sigma_g1s\pi_u2p$		0·66	104	0·75
(Werner bands)	1·2	$\sigma_g1s\pi_u2p$		0·34	104	0·38
	1		$1s2p\pi$	0·65	104	0·73
	1		$1s2p\pi$	(0·32)	(110)	(0·38)
	1·2		$1s2p\pi$	0·33	104	0·37
$^1\Sigma^+_u$	1		$1s3p\sigma$	(0·036)	(98)	(0·038)
$^1\Pi_u$	1		$1s3p\pi$	(0·049)	(98)	(0·052)
$^1\Sigma^+$	1		$1s3s$	(0·008)	(98)	(0·009)

(*a*) Effective nuclear charge for $1s$ AO's ; $Z=1$ for all other AO's.

(*b*) Of upper state ; lower state $(\sigma_g1s)^2$, $^1\Sigma^+_g$ (equation (16)) with $Z(1s)=1$ or 1·2.

(*c*) Of upper state ; lower state $1s1s$, $^1\Sigma^+_g$ (equation (16)), $Z(1s)=1$ or 1·2.

(*d*) Values in parentheses are Mrowka's assumed ν values, and D and f values we obtain from Mrowka's A_iM_i values (l.c. p. 306) assuming $A_iM_i=\bar{U}_{00}Q^2_{iN}=2(1+S^2_{1s})Q^2_{iN}$ and $D_{iN}=G_iQ^2_{iN}$. This relation neglects only the lack of orthogonality among some of the wave functions used, which Mrowka took into account in obtaining his A_i's.

(*e*) AO computations were made by Mulliken (1939 b) with $Z=0{\cdot}725$ for the upper state ; f was then 0·56.

We may conclude from equation (26) that an upper limit for the f value of the Lyman bands is 0·84. The computed values of 0·68 (LCAO) and 0·49 (AO) approach this limit so closely as to leave little margin for the Werner and other bands. To examine this point further, we have computed f values of the Werner bands (cf. table 1), with the result that the sum of the f values, Lyman plus Werner, considerably exceeds 0·84: namely, 1·43 (LCAO) or 1·23 (AO).

Before discussing this discrepancy, we shall describe the computations on the Werner bands. The transition is the same as that for Li_2 described by equation (14), except that $1s$ and $\sigma_g 1s$ replace $2s$ and $\sigma_g 2s$. The upper state is a $^1\Pi_u$ state of electron configuration $1s\,2p\pi$ (AO) or $(\sigma_g 1s)(\pi_u 2p)$(MO). The $\pi_u 2p$ MO can be written in LCAO form as $(2p\pi_A + 2p\pi_B)/\sqrt{2(1+S_\pi)}$, but because of the large size of the $2p\pi$AO's as compared with the internuclear distance in state N, S_π is nearly one and the LCAO form is nearly identical with that of a $2p\pi$ hydrogen AO centred around the mid-point of the molecule. In short, the orbital is essentially a Rydberg MO. The computation by the LCAO method is closely analogous to that for the V, N transition, but with $G=2$ in equation (1).

In the AO approximation also, the computation is straightforward once the wave functions are written down. For the N state the wave function is given by equation (18); for the $^1\Pi_u$ state it is

$$\psi \sim \frac{[1s_A(1)2p\pi_B(2) + 1s_A(2)2p\pi_B(1) + 1s_B(1)2p\pi_A(2) + 1s_B(2)2p\pi_A(1)]}{2\sqrt{1+S_{1s}S_{2p\pi}}}.$$

As we have just shown, the Lyman plus Werner f value computed by either the AO or the LCAO method exceeds the observed value. In seeking an improvement, we may compare the hydrogen molecule with the helium atom. If in H_2 we could squeeze the nuclei together, the $\sigma_g 1s$, $\sigma_u 1s$, and $\pi_u 2p$ MO's would respectively become $1s$, $2p\sigma$, and $2p\pi$ of helium, and the Lyman and Werner transitions together would go over into the transition $1s2p$, $^1P \leftarrow 1s^2$, ^1S of the helium atom. For the latter transition a theoretical f value of 0·27 (good to 30 % or better) has been computed (Wheeler, 1933); of this total, 0·09 would correspond to the Lyman bands and 0·18 to the Werner.

The foregoing comparison of our MO wave functions with helium atom functions suggests two ways of improving the former:

(i) It is to be noted that in helium, the $1s$ AO corresponds to an effective nuclear charge Z greater than one. For $2p\sigma$ and $2p\pi$, however, Z is still practically one. Hence it should be an improvement to increase Z for the $1s$ hydrogen AO's in our LCAO form for $\sigma_g 1s$ from the value 1 so far used to something greater. We have therefore made new computations using for $\sigma_g 1s$, $Z(1s)=1{\cdot}2$, a value found by the variation method (Coulson, 1937), to give about the best possible LCAO value of the energy of the normal state of hydrogen. For $\pi_u 2p$, we have still assumed $Z=1$, and used the same LCAO form as before.

(ii) To improve $\sigma_u 1s$, we might again alter Z; here it should be decreased (Coulson, 1937; Mulliken, 1940b). It appeared probable, however, that a still greater improvement would be effected by going over to the Rydberg form $2p\sigma_u$. As in the case of $2p\pi_u$, the Rydberg form differs very little from a convenient LCAO form—here $(2p\sigma_A + 2p\sigma_B)/\sqrt{2(1+S_{2p\sigma})}$—which we have therefore used. That the MO configuration $\sigma_g 1s\,\sigma_u 2p$ gives a somewhat better approximation to the V state of hydrogen than $\sigma_g 1s\,\sigma_u 1s$ is indicated by a direct comparison (cf. table 1) of intensities computed for the two cases using $Z(1s)=1$. Making the further change to $Z(1s)=1{\cdot}2$, the computed Lyman-Werner f sum by MO's is now 0·62, as compared with an *observed* Lyman-Werner f of about 0·65, after making allowance for the higher frequency band systems.

The comparison is therefore as satisfactory as can be expected until we have a better observed f value.*

The value $Z(1s)=1{\cdot}2$ has also been found (Coulson, 1937) to improve the AO computed energy for the normal state, and we have accordingly made AO computations paralleling the improved MO computations. But now, for state V, to obtain an improvement on the previously used $1s^2(H^+H^-)$ wave functions is less obvious than in the MO case. We have tried, however (both for $Z(1s)=1$ and $Z(1s)=1{\cdot}2$), the $^1\Sigma^+_u$ function of the configuration $1s2p\sigma$; this is found (see table 1) to give results similar to those for the MO configuration $\sigma_g 1s\,\sigma_u 2p$. The wave function is similar to the AO function given above for the upper state of the Werner bands, except that $2p\sigma$ replaces $2p\pi$, and the denominator is $2(1+S_{1s,2p\sigma}+S^2_{1s,2p\sigma})^{1/2}$. The computed Lyman-Werner f sum is now 0·58, close to the observed value. We have also included in table 1 computations for the $^1\Sigma^+_u$ function of the configuration $1s2s$, but do not believe that it is a good approximation for the V state.

For the V states of molecules other than hydrogen, including all those discussed in the following sections of this *Report*, the excited MO's of the V states are less Rydberg-like than in hydrogen, and this indicates that the MO approximation analogous to $\sigma_g 1s\,\sigma_u 1s$ for hydrogen, and the AO ionic approximation analogous to $1s^2$ for hydrogen with suitably adjusted AO Z's, are without much doubt the best simple approximations we can find.†

(c) *Computations on dispersion and polarizability of hydrogen*

The first quantum-mechanical calculations of D and f values of molecular electronic transitions were apparently those of Mrowka (1932, 33). They were

* Still better computations could be made with really accurate wave functions of the James-Coolidge type; these are better than the best AO and LCAO forms. They have been determined for states N and V for their respective equilibrium distances (James and Coolidge, 1933, 39; James, 1934, 35), but this is not just what is needed for the absorption spectrum; instead we require functions for all states at the equilibrium inter-nuclear distance of state N. James and Coolidge have, however, made absolute intensity computations for the well-known continuous emission spectrum of the hydrogen molecule; this work will be considered in § 9.

† There are some molecules, however, particularly those like Li_2, in which it is fairly certain that the V state wave function cannot be approximated even roughly by AO wave functions of ionic type.

made primarily with the object of computing theoretically the polarizability, and also the dispersion, of hydrogen gas. Mrowka used exclusively AO approximation wave functions with (presumably) $Z=1$. His results are listed in parentheses in table 1.* Mrowka was apparently successful in computing the polarizability, but it was later found (cf. Stuart, 1936; Mrowka, 1939) that there were errors in his calculations.

The polarizability of a molecule in an electric field directed along its q axis is given theoretically by

$$\alpha_q=(3e^2/4\pi^2mc^2)\Sigma_j^{(q)}f_j/G_j\nu_j^2), \qquad \ldots\ldots(27)$$

the summation being taken only over those ν's for which the transition is polarized along the q axis. The *mean* polarizability α per molecule, for molecules randomly oriented in an electric field, is given by

$$\alpha=(\alpha_x+\alpha_y+\alpha_z)/3=(e^2/4\pi^2mc^2)\Sigma_j f_j/G\nu_j^2. \qquad \ldots\ldots(28)$$

In the present case, transitions from the normal to the $^1\Sigma^+_u$ upper states are z-polarized, and those to the $^1\Pi_u$ states are all half x- and half y-polarized. Mrowka's computations were on the states $1sns$, $1snp\sigma$, and $1snp\pi$, with $n=2$ and 3, and he estimated the contributions by states with higher n. He obtained $\alpha_z=6{\cdot}14\times10^{-25}$, $\alpha_x=\alpha_y=8{\cdot}49\times10^{-25}$ cc. The mean α of $7{\cdot}71\times10^{-25}$ cc. was in good agreement with the experimental value. The latter, $\alpha=7{\cdot}95\times10^{-25}$, is obtained by using the constants of equation (26) in equation (28). His computation also gave for the anisotropy of the polarizability the value $\alpha_x-\alpha_z=2{\cdot}35\times10^{-25}$ cc., while the value $+2{\cdot}2\times10^{-25}$ has been obtained experimentally (Stuart, 1936; Mrowka, 1939) from the degree of depolarization of molecular light scattering in hydrogen.

Although Mrowka has indicated (Stuart, 1936; Mrowka, 1939) that he has found errors in his computations, we are unable to find any statement of corrections, except that he now gets $\alpha_z>\alpha_x$. Therefore, we shall point out here certain discrepancies between his published results and our computations. The inclusion in equation (27) of terms corresponding to photo-ionization, which Mrowka ignored in his computations, would considerably increase his computed value of α. As we have seen in § 5 (b), these photo-ionization terms correspond approximately to the second term in equation (26), and would account for about one-third of α. In addition, there are further difficulties. Although Mrowka gives the values $\alpha_x=\alpha_y=8{\cdot}49\times10^{-25}$ cc., these seem to be twice the values which actually correspond to his computed A_iM_i's; the value he gives for α_z, however, is consistent with his A_iM_i's. The doubled values for α_x and α_y fortuitously compensate approximately for the absence of the photo-ionization terms. Moreover, his conclusion that $\alpha_x>\alpha_z$ was based on the doubled values, and is now reversed.

Further, in three cases ($1s2s$, $1s2p\sigma$, and $1s2p\pi$), where we have made AO computations with $Z=1$, we get D and f values about twice Mrowka's, and we

* Mrowka gives only quantities which he calls $A_{0i}M_{0i}$, in units of a_0^2 (radius of 1-quantum Bohr orbit of hydrogen); these are essentially proportional to Q^2 values, and from them we obtain D and f values in the manner described in footnote *d* of table 1.

have carefully checked our calculations. This discrepancy is too large to be accounted for by differences in our assumptions, such as the slightly different value used for the inter-nuclear distance, or our neglect of the lack of exact orthogonality among some of the wave functions used. Our conclusion is that the final revised values of Mrowka's computed α would be nearly twice those observed. This, in fact, is what would be expected for $Z=1$ according to our computations as discussed above. Good agreement has been obtained, however (see below), using $Z=1{\cdot}2$ for the normal state, with the improved LCAO excited orbitals described above for the various $^1\Sigma^+_u$ and $^1\Pi_u$ states. Apparently fortuitously, Mrowka's D's and α agree approximately with our AO results for $Z=1{\cdot}2$.

Since Mrowka's original work, several writers (Steensholt, 1935; Hirschfelder, 1935; Easthope, 1936) have made approximate computations by other methods on the polarizability of hydrogen and its anisotropy. Although probably none of these is very exact, they all agree in making $\alpha_z > \alpha_x$.

The foregoing discussion illustrates the possibility of an understanding and approximate computation of at least the lowest-frequency terms in dispersion and polarizability formulas, and in expressions governing optical anisotropy. These and the related topic of molecular refractivities will be further considered in § 8.

Since the foregoing paragraphs were written we have extended our own computations on absorption strengths so as to obtain definite values of the polarizability and anisotropy. The new computations were made by the MO method and include transitions to all discrete and also continuous levels with one electron excited to an orbital of the type $np\sigma$ or $np\pi$. We have computed directly the f values for transitions from the ground state to states with $n=2$, 3, and 4, in the way described above for $n=2$. All computations were made with $Z(1s)=1{\cdot}2$ and $Z(np\sigma \text{ or } np\pi)=1$. While the Z for the $1s$ electron in the excited state wave function should actually be about 1·39 (Coulson, 1937), our computations show that using the latter value instead of 1·2 would make a difference in Q of only about $\frac{1}{2}$ per cent.

The Q for each transition comes out in our computations as the sum of two parts: one part is an ordinary H atom Q integral (except for the change to $Z(1s)=1{\cdot}2$), just as if for a jump of the electron between two AO's on the same nucleus; the second part is just as if for a corresponding jump from nucleus A to nucleus B. We have used the computations for $n=2$, 3, and 4 to estimate for higher n values the ratio of the second of these integrals to the first, and thus have been able to obtain f values for transitions to the higher discrete molecular states by using the known H-atom integrals, corrected to $Z(1s)=1{\cdot}2$. The frequencies of the transitions were obtained from Rydberg series formulae for the higher molecular states, using the frequencies of the Lyman and Werner bands to determine quantum defects for $2p\sigma$ and $2p\pi$. The distribution of f with frequency for the continuous states was determined in a similar way, making use of the results obtained by Sugiura (1927) for the H atom.

The computed total Σf resulting from all transitions from the ground state now comes to 1·92. The agreement with the rigorous theoretical value of 2 is quite satisfactory. Since our computations give all the discrete and continuum f's as a function of ν, we may now compute the polarizability directly using equation (27), and find $\alpha_z = 9{\cdot}2 \times 10^{-25}$ cc., $a_x = a_y = 7{\cdot}0 \times 10^{-25}$ cc. The computed anisotropy is then $\alpha_z - \alpha_x = 2{\cdot}2 \times 10^{-25}$ cc., and the mean polarizability $7{\cdot}7 \times 10^{-25}$ cc.

The slight deficiency in Σf and in α can be explained in part by excited states (including doubly excited states) other than those which we have considered. The results, although not conclusive in this respect, encourage the idea that the MO wave functions used are rather surprisingly good approximations.*

§ 6. TRANSITION PROBABILITIES IN OTHER MOLECULES

(a) *Methods of computation*

The evaluation of Q of equation (1) is always analogous to that for $Q_{V,N}$ in hydrogen, but is in general more complicated, largely because the complete wave functions of many electron molecules are more complicated than those of a two-electron molecule. A helpful simplification can, however, be made without appreciable error, and is applicable to any molecule, diatomic or polyatomic: all those electrons of the electron configuration which remain in unbroken closed shells during the transition may be ignored, and only as many electrons as are present in shells which are broken or formed during the transition need be considered (Mulliken, 1940 b).

Very often this reduces the computation of Q to a two-electron problem. When this happens in the MO approximation, the transition is of the form $(\chi_k)^2 \rightarrow \chi_k \chi_l$, or vice versa, where χ_k and χ_l represent two MO's. Here, by a series of steps exactly analogous to those used above for $Q_{V,N}$ of H_2 (cf. equations (19)–(21)), it is easily shown that

$$Q(\mathrm{MO}) = \sqrt{2} \int \chi_k r \chi_l d\tau, \qquad \ldots\ldots (29)$$

where the integration is over the space of a single electron. By substituting suitable LCAO forms for χ_k and χ_l, Q can be evaluated in LCAO approximation. Other cases must be considered individually, but in the MO approximation all lead to an equation which differs from equation (29) only by a constant factor.

The problem then reduces to that of finding suitable LCAO forms for the MO's in equation (29). For many symmetrical diatomic molecules these are of simple forms like σ_g and σ_u of hydrogen, but in some molecules such as C_2 and N_2 there is the complication of more or less mixing of $ns\sigma$ and $np\sigma$ AO's in the LCAO forms. The amount of s, p mixing often has a large influence on the value of Q, but unfortunately is not easy to determine; the writers are now investigating this matter.

* *Note added in proof*: Since these four additional paragraphs were written the authors have discovered that their estimates of f for the continuum are less certain than was first thought.

In unsymmetrical diatomic molecules, another complication occurs, namely, the presence of unequal LCAO coefficients (e.g., $a \neq b$ in equation (11)), whose determination presents an additional problem. If the molecules are not highly polar, however, it is often not a bad approximation in such cases to assume equality of the coefficients in question (Mulliken, 1940 e). For highly polar molecules, the coefficients may be determined by solution of a secular equation which includes explicitly the difference in electro-negativity of the two atoms.

In polyatomic molecules, the AO method is unwieldy, and most of the computations have been made with the LCAO MO approximation. The determination of the LCAO forms usually requires a special investigation; this problem has been fairly satisfactorily solved for certain types of MO, including those involved in the important class of V, N transitions in double-bonded organic molecules (discussed in § 7).

In V, N transitions, in LCAO approximation, the value of Q is found not to depend on the particular AO's, ϕ_i, appearing in the LCAO MO expressions, since the evaluation leads only to integrals such as $\phi_i^2 z\, dv$, which are independent of the form of ϕ_i (cf., e.g., equation (22)). For Q, N and other transitions, however, one finds integrals such as $\int \phi_i x \phi_j dv$, whose evaluation requires a knowledge of the explicit forms of the actual ϕ's. Since the latter are known only approximately, this introduces an error not present in the LCAO computation of V, N transition probabilities. Similar remarks apply to AO computations, except that here some dependence on the assumed forms of the AO's is present even for V, N transitions. However, this dependence enters only through the S factor in equations such as (24); this plays a more prominent rôle in the AO than in the MO computed intensities.

In general, the intensities computed from the AO and MO approximations differ more than was the case for H_2, because S in other cases usually is numerically smaller (cf. equations (23) and (24), and table 2). Since the true wave functions are believed to be intermediate in type between those of the AO and MO approximations, it seems plausible in making comparison with experiment to expect experimental intensities to fall between those computed by the two approximations —provided, of course, that the electron configurations in any given case correspond to those involved in the actual transition, and provided suitable Z values have been used.

In discussing a set of analogous transitions in different molecules, it is often possible to determine an " empirical correction factor " by comparing computed with observed values in well-established cases. This factor may then be used in making predictions for other similar cases where good experimental values are not available.

(b) *V, N transitions in table* 2

The D and f values for V, N transitions given in table 2 have been obtained like those in § 5 (a) for hydrogen, using equations (23) and (24). The transitions all

Table 2. Computed S values and transition probabilities for V, N transitions in simple molecules (cf. Mulliken, 1939 b)

Molecule	State N				State V		Computed			Observed $\nu_{max} \times 10^{-3}$ (vapour)	Semi-theoretical[e] f
	Species	r (A.)	MO excited	Calculated S	Species	Excited MO	Q (A.)[a] LCAO	Q (A.)[a] AO	D (A.)[b]		
H_2[d]	$^1\Sigma^+_g$	0·74	σ_g	0·75	$^1\Sigma^+_u$	σ_u	0·80	0·68	0·54	99[d]	0·58
F_2	$^1\Sigma^+_g$	1·435[e]	σ_g	0·17	$^1\Sigma^+_u$	σ_u	1·03	0·25	0·41	—	—
Cl_2	$^1\Sigma^+_g$	1·98	σ_g	0·30	$^1\Sigma^+_u$	σ_u	1·46	0·60	1·06	(72)[g]	0·83
Br_2	$^1\Sigma^+_g$	2·28	σ_g	(0·31)	$^1\Sigma^+_u$	σ_u	1·69	0·71	1·44	(64)[g]	1·00
I_2[f]	$^1\Sigma^+_g$	2·66	σ_g	0·31	$^1\Sigma^+_u$	σ_u	1·98	0·84	1·99	56	1·21
O_2	$^3\Sigma^-_g$	1·20	δ_g	0·32	$^3\Sigma^-_u$	σ_u	0·90	0·39	0·41	(100)[g]	0·44
O_2	$^3\Sigma^-_g$	1·20	π_u	0·17	$^3\Sigma^-_u$	π_g	0·86	0·20	0·28	69[h]	0·21
C_2H_4	$^1A_{1g}$	1·33	b_{3u}	0·27	$^1B_{1u}$	b_{2g}	0·98	0·36	0·45	61[i]	0·30
HI	$^1\Sigma^+$	1·60	σ	0·50	$^1\Sigma^+$	σ^*	1·31	0·83	1·14	(78)[g]	0·96[j]

(*a*) LCAO MO by equation (23) ; AO by equation (24).
(*b*) By equation (1) from mean of AO and LCAO computed Q's.
(*c*) By equation (2) using the theoretical D and observed ν_{max}.
(*d*) Lyman bands ; for other computations see table 1. The experimental f for Lyman plus Werner bands is discussed in § 5 (b).
(*e*) $r_0 = 1{\cdot}435 \pm 0{\cdot}01$ from unpublished data of Rogers, Schomaker, and Stevenson.
(*f*) D bands ; see note (*g*) of table 3.
(*g*) Estimated ; spectrum not yet observed.
(*h*) Schumann continuum ; observed $f = 0{\cdot}193$.
(*i*) Strong bands or continuum.
(*j*) Assumed non-polar.

occur at rather high frequencies, and consequently little is known experimentally about their intensities. In the one case (oxygen) for which we have good experimental data, the observed f value agrees well with that determined from the average Q of the MO and AO computations.

(c) *Q, N transitions in table 3; some transitions of astrophysical interest*

Table 3 contains similar computations on D and f values for Q, N transitions (Mulliken, 1940 b, 40 c; McMurry, 1941 a). Here the absolute agreement between computed and observed values is not very satisfactory, the observed being smaller than both the computed AO and the computed LCAO values. However, the relative values for different molecules come out for the most part in agreement with the computations. Particularly satisfactory is the series F_2, Cl_2, Br_2 (I_2 and HI do not fit in, but there are probably special reasons why they are abnormal (Mulliken, 1940 b).

The poorer absolute agreements here than for the V, N transitions may be attributed chiefly to the fact noted above that the Q, N computations are especially sensitive to the forms and sizes of the MO's or AO's involved. In the computations, modified hydrogen-like AO's of the analytical forms and with the effective nuclear charges Z suggested by Slater for uncombined atoms were used, both in the AO computations and for the LCAO forms in the MO computations. In the cases of F_2, Cl_2, and Br_2, we find that an increase of the Slater Z's by 10–15% is sufficient to give a satisfactory agreement of the same kind as for V, N transitions.*

An additional way in which the wave functions might (in fact, *ought to*) be improved is by the use of polarized AO's. In the case of $p\sigma$ AO's, polarization can in part be represented by the admixture of some $s\sigma$. Recent unpublished computations we have made on the Q, N transition in OH indicate that s, $p\sigma$ mixing there greatly reduces the computed f value, bringing it near to the experimental value. Preliminary computations on the corresponding transitions in CH and NH, based on analogous determinations of s, $p\sigma$ mixing, give considerably larger f values than for OH; more definite results will be published in a later paper. The violet CN bands are also expected to be sensitive to s, $p\sigma$ mixing; the electron configurations concerned are very similar to those for N_2^+ given in equation (15). In all of these the effect of the s, $p\sigma$ mixing may be very pronounced.

The transitions just mentioned are of especial interest since they occur strongly in the spectra of astrophysical sources. Thus a reliable computation of their f values would lead to a great improvement in the determination of the abundances of O, N, and C in the sun, stars and in interstellar space. Some results on the problem have just been published by Lyddane, Rogers and Roach (1941), who give ***relative*** f values for CH, NH, and OH, and for C_2 and CN. No details of the computations have been given, and it is uncertain to what extent s, $p\sigma$ mixing has been considered.

* The poor agreement using the Slater Z's results from any unusually great sensitivity of the D values to the Z's (cf. § 9). Arguments given in Mulliken (1940 b) against the use of increased Z's appear inconclusive on reconsideration.

Table 3. Transition probabilities for Q, N transitions[a]

Molecule	State N		State Q	Q, N transition[b] D (A.U.) computed		Observed D (A.U.)	Nature of spectrum
	Species	r_e (A.)	Species	LCAO	AO		
F_2	$^1\Sigma^+_g$	1·435[c]	$^{3,1}\Pi_u$	0·0060	0·00058	0·00068	Continuum
Cl_2	$^1\Sigma^+_g$	1·983	$^{3,1}\Pi_u$	0·063	0·014	0·0056	Continuum
Br_2	$^1\Sigma^+_g$	2·284	$^{3,1}\Pi_u$	(0·084)	(0·024)	0·012–0·014	Continuum and bands
I_2	$^1\Sigma^+_g$	2·667	$^{3,1}\Pi_u$	0·094[g]	0·024[g]	(0·0012)	Continuum and bands
HF[d]	$^1\Sigma^+$	0·917	$^{3,1}\Pi$	0·078	0·020		Continuum
HCl[d]	$^1\Sigma^+$	1·275	$^{3,1}\Pi$	0·18	0·088		Continuum
HBr[d]	$^1\Sigma^+$	1·414	$^{3,1}\Pi$	(0·19)	(0·098)	0·058	Continuum
HI[d]	$^1\Sigma^+$	1·604	$^{3,1}\Pi$	0·20	0·11	not $>$0·02	Continuum
CH_3Cl[d]	$^1A_1$	1·77 (C−Cl)	$^{3,1}E$	0·044[e] 0·032 0·089	0·012[e] 0·0084 0·038	(0·01 ?)	Continuum
H_2CO	$^1A_1$	1·22 (C=O)	$^{3,1}B_2$	0·08	0·06		
OH[d]	$^2\Pi$	0·971	$^2\Sigma^+$	0·012[f]	0·0017[f]	0·0008	Bands

(*a*) The intensities for H_2CO were computed by McMurry (1941 a); for all other molecules by Mulliken (Mulliken, 1940 b, 40 c). Except for H_2CO and OH, D values given here are all twice those in references quoted, because of the change in definition; cf. equation (1) and note.

(*b*) The nature of the transition is described for the halogens by equations (12) and (13) in the text, for the hydrogen halides by (8) and (9).

(*c*) See note (*e*) of table 2.

(*d*) The computations were made assuming the bond in state N to be non-polar; if the bond in HF (the most polar case) is assumed 50 % polar, D is 0·105 by the LCAO MO method, 0·084 by the AO (Mulliken, 1940 c).

(*e*) The Q, N transition in CH_3Cl is analogous to that for the hydrogen halides (cf. equations (8) and (9)), with the complication that the σ and σ^* orbitals are formed from $3p\sigma_{Cl}$, combined with $2p\sigma_C$ or $2s_C$ or a *mixture* of these. The three sets of values for D given in the table correspond to the three assumptions just given; the particular *mixture* assumed here was $(s_C + \sqrt{3}p\sigma_C)/2$.

(*f*) See equations (10) and (11) for OH. A preliminary determination of the amount of s, p mixing in OH indicates that D by the LCAO MO computations using improved MO's is much less than that given here.

(*g*) The computed D values for I_2 have been corrected for a numerical error in S in Mulliken, 1940 b. The error is due to a wrong coefficient in equation (22) of that paper; the term $23\gamma^7/66150$ should read $13\gamma^7/66150$.

An example of the use of computed intensities in deciding between alternative interpretations of a spectrum is provided by McMurry's (1941 a) computations on the Q, N transition in aldehydes and ketones. Mulliken (1935 b; McMurry and Mulliken, 1940) showed that the weak transition near $\lambda 2900$ must be identified either as (i) a certain forbidden transition, or (ii) an allowed transition of the Q, N type. McMurry's computed intensities for the Q, N transition are much too large to be compatible with identification of the $\lambda 2900$ absorption as Q, N, and it may therefore be identified with the forbidden transition. On the other hand, a stronger absorption region observed near $\lambda 1900$ may probably be identified as Q, N (McMurry, 1941 a).

(d) *Monosubstituted benzenes in table* 4

Table 4 gives intensities for some monosubstituted benzenes, computed by Sklar (1939) essentially by the LCAO MO method. In benzene, the longest wave-length transition is forbidden by the electronic selection rules on account of the high symmetry of the molecule, but actually occurs, though weakly, because of interaction with vibrations (cf. § 9). In the substituted benzenes, the six-fold symmetry is disturbed, and the transition is allowed by the selection rules, and

Table 4. Transition probabilities for $\lambda 2700$ band of substituted bezenes[a]

Molecule	Theoretical f	Observed f
$C_6H_5CH_3$	0·0005	0·0013
$C_6H_5NH_2$	0·0140	0·0246
C_6H_5OH	0·0060	0·0197
C_6H_5F	0·0003	(0·0010)[b]

(*a*) From Sklar's table 1 (Sklar, 1939), except that his experimental f is corrected by a factor of 2·3/1·4. The 2·3 arises because of Sklar's erroneous identification of $\epsilon_v c$ with k_v (cf. § 2 (a)). The factor of 1·4 is because Sklar included the index of refraction of the solution in determining f; while this procedure is theoretically correct in applying the Lorentz-Lorenz correction, our conclusion in § 2 (b) is that, on empirical grounds, no correction should be applied.

(*b*) The experimental data are for C_6H_5Cl.

is expected to occur with more or less intensity, depending on the nature of the substituent. The problem of computing intensities is greatly complicated here by the necessity of including atoms other than carbon. Nevertheless, Sklar succeeded in computing intensities of the right order of magnitude (about 1/3 or 1/2 as large as the observed), and differing correctly for different substituents. For details concerning the computations, the reader is referred to Sklar's paper.

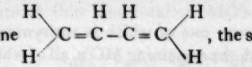
§ 7. *V, N* TRANSITIONS IN DOUBLE-BONDED ORGANIC MOLECULES*

(a) *Characteristics of orbitals in conjugated bonds*

In many of the complicated molecules of organic chemistry, V, N transitions arising from the transfer of a double-bond electron from a bonding to a

* See Mulliken, 1939 c–39 h.

corresponding anti-bonding MO are of especial interest and importance, since they are responsible for colour in many pigments and dyes (cf. Förster, 1939).

As is well known, an ordinary valence bond involves two electrons shared between two participating atoms; for a double bond there are four shared electrons, and a triple bond six. If two double bonds are separated by a single bond, as in butadiene  , the system no longer behaves quite as one might expect for two ordinary isolated double bonds and one single bond, and the double bonds are referred to as " conjugated ". Conjugation results in a partial acquisition of double bond character by the single bonds and vice versa, and is explainable by quantum-mechanical resonance effects. The term " conjugated " is applicable to any system of alternating multiple and single bonds.

The orbitals occupied by the electrons forming a bond may be divided into two types called σ and π, which are respectively symmetrical and anti-symmetrical in a plane passing through the axis of the bond. An ordinary single bond consists of a pair of σ electrons; a double bond contains an additional pair of π electrons, and a triple bond two pairs of π electrons.

In the presence of an axis of symmetry, as in the case of a triple bond, there are two independent but equivalent π orbitals (say π_x and π_y) differing from each other only by a rotation of 90° around the axis. It is known from chemical and spectroscopic data that π electrons are more loosely bound than σ electrons, and are responsible for the chemically " unsaturated " characteristics of molecules containing double or triple bonds. Hence these electrons can appropriately be referred to as " unsaturation electrons " (Mulliken, 1939 c); they are also called " mobile electrons " (Lennard-Jones, 1937).

The outstanding characteristics of the unsaturation electrons in double bonds is that they enforce a plane arrangement of all atoms directly connected to the double-bonded atoms, or involved in a system of conjugated bonds. In our computations, this plane will be taken as the yz plane, and the unsaturation orbitals, which are antisymmetrical to this plane, are π_x orbitals. Here π_x and π_y are no longer equivalent. The double bond consists of a pair of σ and a pair of π_x (or more briefly, x) electrons.

In the discussion thus far we have tacitly assumed that the orbitals are *localized*; that is, that for each chemical bond there is one MO, occupied by two electrons (MO approximation), *or* one pair of AO's each occupied by an electron (AO approximation). The best use of the MO approximation for molecules with conjugated double bonds, however, involves *non-localized* MO's for the unsaturation electrons. These may be constructed as linear combinations of localized MO's and therefore, if the LCAO approximation is used, as linear combinations of AO's. In the case of conjugated double bonds between atoms such as carbon or nitrogen, the AO's are $2p_x$. The non-localized MO's are

related to the localized MO's very much in the same way that normal vibrations of a complex molecule are related to localized vibrations of separate atom-pairs. In both cases, the solution of the non-localized problem involves setting up a secular equation whose roots yield, in the one case the energies of the non-localized LCAO MO's, in the other the frequency parameters of the normal modes of vibration.

The unsaturation MO's, every one of which is antisymmetrical to the plane of the molecule, do not mix with the remaining MO's, all of which are symmetrical to the plane. Thus for the following computations, the electrons in symmetrical orbitals may be neglected. A thorough discussion of molecular orbitals in unsaturated molecules is given in a review article by Hückel (1937).

(b) *V, N transition in ethylene*

If there is just one double bond, as in ethylene $\begin{matrix}H\\H\end{matrix}\!>\!C=C\!<\!\begin{matrix}H\\H\end{matrix}$, there are just two unsaturation electrons. With respect to these (cf. equations (16) and (17) for hydrogen) the MO electron configurations, omitting inner electrons, are

$$N\colon\ \ldots\ (\chi_g)^2,\ {}^1A_{1g},$$
$$V\colon\ \ldots\ (\chi_g)(\chi_u),\ {}^1B_{1u},$$

where, in LCAO approximation,

$$\chi_g=(\phi_1+\phi_2)/\sqrt{2(1+S)},$$
$$\chi_u=(\phi_1-\phi_2)/\sqrt{2(1-S)},$$

which are respectively bonding and antibonding. The difference in the state species symbols between ethylene and hydrogen is, of course, a result of the different symmetries, but is of no importance for the intensity calculations. The non-specific symbols χ for MO's, ϕ for AO's, introduced here, will be used also in the subsequent discussion of more complicated molecules. Here ϕ_1 and ϕ_2 refer to $2p_x$ AO's of carbon atoms 1 and 2. The evaluation of $Q_{V,N}$ now exactly parallels that for hydrogen (§ 5 (a)), and the LCAO approximation leads to

$$Q_{V,N}=\sqrt{2}\int\chi_g\chi_u\,dv=r/\sqrt{2(1-S^2)}. \qquad \ldots\ldots(30)$$

The results of quantitative computations based on equation (30) and on the corresponding AO equation (which can be obtained in a manner analogous to equation (24) for hydrogen) are given in table 2.

(c) *V, N transitions in butadiene, and the effect of molecular shape on intensities in complex molecules*

In discussing molecules with systems of conjugated double bonds, we shall for the sake of simplicity use exclusively the LCAO MO approximation. In such molecules there are several different unsaturation MO's, normally occupied,

together with about an equal number, normally unoccupied, to which an electron may be excited. Thus we have several V states and several V, N transitions. The computations for butadiene in the following paragraphs can easily be extended to more complex molecules.

We have available a $2p_x$ AO for each of the carbon atoms—four in butadiene. We can then construct four unsaturation MO's, which in LCAO form are represented by

$$\chi_i = a_{1i}\phi_1 + a_{2i}\phi_2 + a_{3i}\phi_3 + a_{4i}\phi_4, \qquad \ldots\ldots(31)$$

where ϕ_m represents a $2p_x$ AO of the mth atom. In the normal state, two electrons, with opposite spin, are allotted to each MO until all available electrons are used up. Thus the two lowest-energy orbitals χ_1 and χ_2 are occupied. The excitation of one electron from either of these to either of the two unoccupied higher-energy MO's χ_3 and χ_4 leads to one of four V states, which we shall label V_1 to V_4. Therefore there are four V, N transitions. The computations indicate that the levels V_1 to V_4 should cover a considerable energy range, with V_1 lower and V_4 higher than the single V state in ethylene. Thus the V_1, N transition is expected to come at longer wave-lengths than V, N in ethylene, in agreement with observation (cf. tables 2 and 3).

The coefficients a_{mi} in equation (31) are determined by solution of a secular equation obtained by quantum-mechanical variation theory or perturbation theory. It is easily shown that the condition for the "best values" of the coefficients is that

$$\sum_{m=1}^{4} a_{mi}(H_{mn} - S_{mn}E_i) = 0, \qquad n = 1, 2, 3, 4, \qquad \ldots\ldots(32)$$

where $H_{mn} = \int \phi_m H \phi_n d\tau$, and $S_{mn} = \int \phi_m \phi_n d\tau$; the S integrals are readily computed approximately by using suitable approximate analytical forms for the AO's, e.g. the Slater functions.

Equation (32) represents a set of four linear simultaneous equations in four independent variables; therefore, for significant solutions, the determinant of the coefficients must equal zero. This leads to the secular equation

$$\begin{vmatrix} H_{11}-S_{11}E & H_{21}-S_{21}E & H_{31}-S_{31}E & H_{41}-S_{41}E \\ H_{12}-S_{12}E & H_{22}-S_{22}E & H_{32}-S_{32}E & H_{42}-S_{42}E \\ H_{13}-S_{13}E & H_{23}-S_{23}E & H_{33}-S_{33}E & H_{43}-S_{43}E \\ H_{14}-S_{14}E & H_{24}-S_{24}E & H_{34}-S_{34}E & H_{44}-S_{44}E \end{vmatrix} = 0. \qquad \ldots\ldots(33)$$

The four solutions E_i of this equation are the energies of the four desired MO's, χ_i. Using these energies, the ratios of the coefficients are obtained from equations (32) and the absolute values from the normalization condition. With the LCAO forms thus determined, we derive the intensity of any V, N transition by substitution in equation (29) of the two MO's between which the electron jumps.

The solution of the secular equation, however, presents some difficulties, and

on this account the LCAO method has been applied to polyatomic molecules in three distinct approximations. In all three it is assumed that the H_{mn}'s, other than those for neighbouring atoms may be neglected and that the energy may be expressed in terms of two integrals which are essentially empirical. These are H_{mm}, the "Coulomb" integral, which we shall call α; and $H_{mn}-S_{mn}\alpha$, the "exchange" integral, which we shall call β_{mn}.* The secular equation then reduces to

$$\begin{vmatrix} x & \beta_{21}+S_{21}x & 0 & 0 \\ \beta_{12}+S_{12}x & x & \beta_{32}+S_{32}x & 0 \\ 0 & \beta_{23}+S_{23}x & x & \beta_{43}+S_{43}x \\ 0 & 0 & \beta_{34}+S_{34}x & x \end{vmatrix} = 0, \quad \ldots\ldots(34)$$

where x denotes $\alpha - E$.

The simplest approximation, due to Hückel (1937), now assumes that all the β_{mn}'s are equal (regardless of variations in the internuclear distance), and neglects the Sx's. The second, by Lennard-Jones (1937), still neglects the Sx's but allows for the variation of β_{mn} with internuclear distance. Finally, the third approximation also includes the Sx's (Mulliken, Rieke and Brown, 1941; Mulliken and Rieke, 1941). While the differences between the three approximations prove to be important for some molecular properties, we have found that they have very little effect on the computed intensities. In the accompanying tables, the latter are based partly on the Hückel, partly on the third procedure.

The result of the integration indicated in equation (29) can be expressed in a very instructive vector form† using the first or the second type of approximation (the third leads to some unimportant extra terms which complicate the vector expression). Using the second type of approximation, one finds that equation (29) leads to

$$\boldsymbol{Q} = \sum_{m=1}^{4} a_{mk} a_{ml} \boldsymbol{r}_m, \quad \ldots\ldots(35)$$

where $\boldsymbol{r}_m$ gives the location of the mth atom. Different $\boldsymbol{r}_m$'s in equation (35) are always found to occur in pairs with equal and opposite coefficients, so that equation (35) can be rewritten in terms of vector distances $\boldsymbol{r}_{mn}(\boldsymbol{r}_{mn} = \boldsymbol{r}_m - \boldsymbol{r}_n)$ between pairs of atoms in the molecule. For the four V, N transitions in butadiene, in order of increasing frequency, the results are

$$\begin{aligned} \boldsymbol{Q}_{V_1,N} &= 0{\cdot}489\boldsymbol{r}_{14} - 0{\cdot}218\boldsymbol{r}_{23}, \\ \boldsymbol{Q}_{V_2,N} &= 0{\cdot}326\boldsymbol{r}_{12} - 0{\cdot}326\boldsymbol{r}_{34}, \\ \boldsymbol{Q}_{V_3,N} &= 0{\cdot}326\boldsymbol{r}_{12} - 0{\cdot}326\boldsymbol{r}_{34}, \\ \boldsymbol{Q}_{V_4,N} &= 0{\cdot}218\boldsymbol{r}_{14} - 0{\cdot}489\boldsymbol{r}_{23}. \end{aligned} \quad \ldots\ldots(36)$$

* In the literature, the integral H_{mn} alone is often referred to as β; however, it can be shown (cf. Mulliken and Rieke, 1941) that the empirical β obtained in practice actually corresponds to β as defined above.

† This vector form was not described in the original publications.

Figure 1 represents the carbon atom skeleton of the two possible configurations for butadiene, *cis* and *trans*.* The internuclear distances and angles are from electron diffraction data (Schomacher and Pauling, 1939). It is readily seen from equations (36) that the *trans* configuration, with r_{14} large and r_{13} parallel to r_{24}, leads for V_1, N to a very large Q directed approximately parallel to the longest axis of the molecule; for V_2, N and V_3, N to $Q=0$; and for V_4, N to a small Q directed approximately perpendicular to the longest axis (cf. table 5 for numerical results on the intensities). For the *cis* configuration, however,

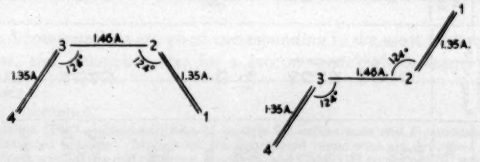


Figure 1. ***Cis*- and *trans*-butadiene.**

r_{14} is much smaller, and r_{13} and r_{24} are no longer parallel; this results in a weaker V_1, N transition polarized parallel to the direction of the single bond, and strong V_2, N and V_3, N polarized perpendicular to the single bond. For either configuration, equations (36) show that the intensities are sensitive to the lengths of the bonds and to the bond angles.

The foregoing results for butadiene serve as an example of the effect of molecular shape on spectral intensities and polarizations in complex molecules. Corresponding effects are also to be expected for other related molecular properties (refractive index, dispersion, polarizability, and so on—cf. §5 and §8).

(d) *Butadiene and substituted butadienes; comparison of theory and experiment*

Observations on the spectrum of butadiene show that the first ultra-violet absorption region, which must be identified with V_1, N, is very intense (cf. table 5), while a second region, presumably V_2, N, is weak (Price and Walsh, 1940) or perhaps absent (Scheibe and Grieneisen, 1934). This indicates, as does other evidence (e.g. Raman data (Kohlrausch and Stockmair, 1935)), that butadiene is predominantly *trans* with perhaps a small admixture of *cis*.

Further evidence is obtained from the spectra of methyl and other derivatives of butadiene, to which in first approximation the preceding discussion of butadiene itself is applicable. In cyclopentadiene and cyclohexadiene, the butadiene bond structure is built into a ring and so necessarily corresponds to the *cis* model of

* The designations *cis* and *trans* here refer to the arrangement around a *single* bond, whereas they are usually used in organic chemistry for the arrangement about a *double* bond. In the latter case, the *cis* and *trans* forms can generally be obtained as separate isomers. In the present case the forms are apparently labile, so that only the more stable form should be found, or a mixture in case both are about equally stable.

Table 5. Transition probabilities for V, N transitions in polyatomic molecules (LCAO MO method)

Molecule	Model[a]	Calculated D (A_0^2) [b, k]				V_1, N transition			
						Observed[c] $\nu_{max} \times 10^{-3}$	Semi-theoretical f		Observed f in solution[f]
		V_1, N	V_2, N	V_3, N	V_4, N		From eq. (3)	Corrected[d]	
Dienes		A. Conjugated polyene chain molecules							
1,3-butadiene	*trans*	2·50	0·00	0·00	0·32	47·7	1·29	0·52	0·53
	cis	1·15	0·62	0·65	0·01		0·60	0·24	
2,3-dimethyl-1,3-butadiene *	*trans*	1·92				45·6	0·95	0·38	0·43
2-methyl-1,3-butadiene *	*trans*	2·49				46·5	1·26	0·50	0·41
	cis	1·05					0·53	0·21	
1-methyl-1,3-butadiene *	*trans*	2·77				46·5	1·40	0·56	(0·5)[g]
1,4-dimethyl-1,3-butadiene *	*trans*	3·06				45·6	1·51	0·60	(0·5)[g]
1,3-cyclopentadiene *	*cis*	0·55	1·07	0·77		43·2	0·26	0·078	0·078[h]
1,3-cyclohexadiene *	*cis*	1·05	0·84	0·83		40·5	0·46	0·138	0·139[h]
Polyenes with even number of carbon atoms in chain									
Octatetraene	I (*trans*)	9·40	0·00(2)	0·15(2)	0·32	(38)	3·88	1·55	
(8-atom chain)	III (*cis*)	5·63	1·72(2)	0·13(2)	0·00		2·32	0·93	
β-carotene	I *trans*)	66·8	0·00(2)			22	17·4	7·0	2·69
(22-atom chain)	II (intermediate)	45·6	0·00(2)	1·97(2)	0·61		11·9	4·8	
Polyenes with odd number of carbon atoms in chain[i]									
$[C_{11}H_{14}]^-$ *	*trans*	16·73	0·00	0·24	0·16				
$C_{11}H_{14}$ *	*trans*	8·37	0·00	0·24	0·08				
Rings		B. Ring and ring-chain molecules							
Benzene, C_6H_6		3·86[j]				56	2·35	0·71	
Styrene,* $C_6H_5-CH=CH_2$		2·42	1·68(2)	0·97	0·40(2)	41	1·13	0·34	0·34
Fulvene,* $C_5H_4=CH_2$		0·22	2·47	0·70	0·00	27	0·06	0·02	*m*
Diphenyl,* $C_6H_5-C_6H_5$		2·95	2·22(4)	1·94(4)	0·00(2)	41	1·38	0·41	0·41
Diphenyl polyenes with even number of carbon atoms (Ph = phenyl = C_6H_5) *									
$Ph-(CH=CH)-Ph$	*trans*	4·98(1)	1·44(4)	0·00(2)	1·94(4)	32	1·84	0·55	0·62
	cis	2·60(1)	1·44(4)			37	1·10	0·33	0·39[m]
$Ph-(CH=CH)_2-Ph$	*trans*	8·11(1)	1·00(6)	0·86(2)		30	2·82	0·85	0·84
	cis	6·77(1)	1·44(4)				2·35	0·71	
$Ph-(CH=CH)_4-Ph$	*trans*	17·26(1)	0·00(2)	0·62(4)	1·16(2)	26	5·24	1·57	1·42
Diphenyl polyenes with odd number of carbon atoms (radicals) *									
$Ph-(CH)-Ph$		4·54(6)	0·58(2)						
$Ph-(CH)_3-Ph$	*trans*	6·74(2)	0·42(4)						

figure 1. In striking agreement with the predictions for this model, the V_1, N transitions in these two molecules are much weaker than for butadiene and its non-cyclic derivatives (cf. table 5).

In the non-cyclic butadienes, the observed V_1, N intensities vary somewhat but are in general agreement with the butadiene *trans* model, with perhaps some admixture of *cis* molecules. The use of space models suggests that the relative stability of *cis* and *trans* forms should depend somewhat on the location of the substituted methyl groups, but only in the case of isoprene (2-methyl—1,3-butadiene) does the space model suggest that the *cis* form should be the more stable.

In table 5 computations are given corresponding to the most probable model in each case, and sometimes also for a second model which seems possible.

NOTES FOR TABLE 5

* Hitherto unpublished.

(*a*) See Mulliken (1939 g) for description of models for octatetraene and β-carotene. Model I is the most elongated possible. Models for the substituted butadienes are described in the text (§ 7). In all non-cyclic dienes and polyenes, C=C−C and C=C−H angles of 120° were assumed. The electron diffraction result of 124° for butadiene would change the calculated D for V_1, N in *trans*-butadiene only to 2·56 A.U.

(*b*) The D's for the dienes were computed by the approximation including S and including hyperconjugation, while for all other molecules in table 5 the simple Hückel approximation was used. The blank spaces indicate values not yet computed. For most of the molecules in table 5, there are more than the four V, N transitions for which results are given here.

(*c*) The observed ν_{max} for the substituted butadienes, benzene, and (dimethyl) fulvene are for vapour ; all others are for solution.

(*d*) The semi-theoretical f is obtained using the calculated D and observed vapour ν_{max} in equation (3) ; solution values of ν_{max} were increased by 2000 $cm.^{-1}$ to obtain estimated vapour ν_{max}.

(*e*) The empirical correction factor used for chain dienes and polyenes is 0·4 ; for ring and ring-chain molecules, 0·3 (cf. § 7 (*g*)). These factors serve to show consistency of agreement within each group of molecules. For very long chains, smaller correction factors seem to be needed (cf. β-carotene).

(*f*) References for experimental data on the dienes are given in Mulliken, 1939 c ; Henri and Pickett, 1939 ; Pickett, Paddock and Sackter, 1941 ; in addition, for the end-substituted dienes see Booker, Evans, and Gillam (1940). For β-carotene see Mulliken, 1939 g. The experimental data for diphenyl and the diphenyl polyenes are from Hausser, Kuhn, Smakula, and Hoffer (1935) except for *cis*-stilbene (Smakula and Wassermann, 1931). For styrene, see Smakula (1934). The measurements of Lewis, Magel and Lipkin (1940) do not permit a satisfactory determination of the f value for *cis*-stilbene ; for *trans*-stilbene, they indicate an f value of 0·67, in good agreement with the earlier results.

(*g*) The observed f value is uncertain ; the observations (note (*f*)) give 0·44 for l-methyl-1,3-butadiene and 0·39 for 1,4-dimethyl—1,3-butadiene, but according to Stark's observations (cf. Mulliken, 1939 c) the f value of the latter is perhaps 3 or 4 times that of 2-methyl—1,3-butadiene.

(*h*) In the vapour state, the observed f is 0·084 for cyclopentadiene and 0·135 for cyclohexadiene.

(*i*) Cf. Mulliken, 1939 h. The last occupied normal state MO is occupied by only one electron in $C_{11}H_{14}$; hence transitions from this MO leading to V states are weaker by 1/2 than the same transitions in $[C_{11}H_{14}]^-$, in which the MO concerned is full. The transition V_2, N is from a lower, fully occupied MO.

(*k*) For the polyenes, diphenyl, and the diphenyl polyenes, many of the V states, in the Hückel approximation used here, are degenerate. In such cases the D given is the sum over the transitions to all the states of equal energy, and the figure in parentheses indicates the number of such states.

(*l*) This is the only allowed V, N transition (cf. Mulliken, 1939 e).

(*m*) According to observations by Stark *et al.* (1913), dimethylfulvene has absorption bands at 27,000, 39,000, and 48,000 $cm.^{-1}$ with the band at 39,000 strongest and 48,000 next in intensity. The computed intensities are seen to vary similarly.

(*n*) The good agreement between computed and observed f values for *cis*-stilbene is probably accidental ; see text.

For cyclopentadiene, the dimensions and bond angles are based on information from electron diffraction (Schomacher and Pauling, 1939); for cyclohexadiene the model represented in figure 2 was assumed. The names of the various substituted non-cyclic butadienes may be understood by reference to figure 1. The numerals preceding " methyl " or " dimethyl " indicate the positions of attachment of the methyl groups, with the carbon atoms of the C=C–C=C skeleton numbered as in figure 1; the numerals preceding "—— diene " refer to the positions of the double bonds.

In making the computations for the substituted butadienes, allowance has been made for a phenomenon called hyperconjugation (Mulliken, Rieke and Brown, 1941; Mulliken and Rieke, 1941). This is a variety of weak conjugation

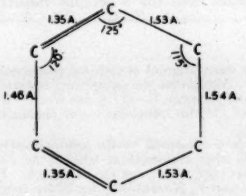


Figure 2. Model assumed for cyclohexadiene.

which exists between multiple bonds and methyl or other groups, and produces effects qualitatively similar to those for additional conjugation bonds, but quantitatively much smaller. The computations indicate, in agreement with the observations, that V_1, N is shifted toward longer wave-lengths by hyperconjugation, and they explain the unusually large red shifts for the cyclic dienes. They also predict considerable changes in intensity in some cases.

The (not hitherto published) computed intensities for the substituted butadienes in table 5 are preliminary, since they depend on a preliminary determination of empirical parameters governing the amount of participation of the methyl groups in the conjugated system. In spite of this uncertainty and that in the nuclear configurations, the computations indicate quite strongly that there should be appreciable, significant differences between the f values of butadiene and its derivatives. The semi-theoretical f values agree with the observed, within the errors of the latter, and further, are successful in explaining differences of molecular refractivity, discussed in §8(a).

(e) *Polyene chains; organic pigments and dyes*

Butadiene is the simplest representative of the class of conjugated polyene chain molecules. In this class the total number of V, N transitions is predicted to be equal to the square of the number of conjugated double bonds. The computations on longer chains predict for the V_1, N transition a steady shift

toward longer wave-lengths (Hückel, 1937) and (for the extreme *trans* models) a steady increase in intensity (Mulliken, 1939g, 39h), with increasing length of the chain. This is in entire agreement with observation. Many well-known plant and animal pigments are derivatives of conjugated polyene chain molecules—for example, the carotenes contain a chain of eleven conjugated double bonds. Here the V_1, N absorption is astonishingly strong and occurs in the blue and violet. Results of the computations for octatetraene (four conjugated double bonds) and carotene are given in table 5.

For pure hydrocarbon chains, stable molecules exist only for an even number of carbon atoms in the conjugated system, two for each double bond (i.e. *even* polyene chains). There is, however, also an important class of molecules, including many dyes, whose simplest prototypes are polyene chain molecules similar to those just discussed except that the number of atoms forming the conjugated chain is odd instead of even. Often the chain forms a positive ion like $(NH_2 \overline{-} CH \cdots CH \overline{-} NH_2)^+$, where the $\overline{-}$ indicates a bond structure most simply described as being a result of resonance between two forms $NH_2^+ = CH \cdots CH - NH_2$ and $NH_2 - CH \cdots CH = NH_2^+$.

The results of computations for an odd polyene chain hydrocarbon containing eleven carbon atoms are given in table 5. Such a molecule is a " free radical " with an odd number of electrons; by adding an electron, it becomes a normal negative ion; or by substituting NH_2 and NH_2^+ for CH_2 at the two ends, it becomes a normal positive ion. The substitution of nitrogen for carbon should affect the intensity but little. It is thus seen from table 5 that such odd polyene chain molecules should have V, N spectra very similar to those of the corresponding even polyenes. Many dye molecules, for example the polycarbocyanine dyes used as sensitizers in infra-red photography, are salts whose positive ion is built on an odd polyene chain with nitrogen at each end, and the strong longest wave-length infra-red absorption of such a molecule is evidently of the V_1, N type. However, the infra-red sensitizing action apparently depends not on absorption by single molecules but on a co-operative action of many molecules.

The spectra of dye molecules of this and similar types have also been discussed by Pauling (1939), who, using the AO approximation and introducing one or two empirical parameters, has succeeded in reproducing observed frequencies and intensities.

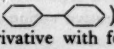
(f) *Ring and ring-chain molecules; steric hindrance*

We have seen from the preceding discussion that the V_1, N transition contains a large part of the total V, N intensity in the case of a molecule of *trans* or elongated structure, but much less for a *cis* or compact structure. The same general tendency appears in the results of computations on a variety of other molecules (table 5). In the very symmetrical six-ring molecule benzene, the longest wave-length transitions are of forbidden type, and there is only one

allowed V, N transition (Mulliken, 1939e), of high intensity, but at about the same frequency as for the unconjugated molecule ethylene. In the less symmetrical five-ring molecule fulvene with conjugation to a double bond attached on one corner, V_1, N is much weaker than higher frequency V, N transitions—in agreement with qualitative experimental observations. In various molecules in which benzene rings are conjugated to double bonds, the intensity of V_1, N (both computed and observed) increases with the maximum overall length of the molecule, becoming very large in the higher diphenyl polyenes.

Stilbene (diphenyl ethylene, $C_6H_5-\underset{C}{\overset{H}{=}}\underset{C}{\overset{H}{=}}-C_6H_5$) is interesting in that intensity measurements have been made on the *cis* and *trans* isomers separately * (Smakula and Wassermann, 1931; Lewis, Magel and Lipkin, 1940). The computed intensities differ exactly as the observed. This is probably somewhat accidental, however, since the use of space models indicates that *cis*-stilbene cannot have all its atoms in one plane, as we have so far assumed for all the ring-chain molecules. If *cis*-stilbene has the plane of one benzene ring perpendicular to that of the other, its V_1, N absorption should resemble that of styrene (phenylethylene); as it happens, the latter's spectrum is almost the same as is predicted for the coplanar model of *cis*-stilbene.

Some interesting examples of the spectra of molecules in which the presence of bulky groups (" steric hindrance ") forces double-bonded groups out of a coplanar arrangement and so breaks up their conjugation have been studied by O'Shaughnessy and Rodebush (1940). Exactly as expected, the spectra revert to those of the constituent internally coplanar groups. Interesting intermediate cases are also found. A fine example of complete loss of conjugation is found in such cases as the substituted diphenyls

. In diphenyl itself () the two rings are coplanar and V_1, N is strong, but in the derivative with four R groups this strong absorption disappears. Other spectra of interest in this connection have been studied by Jones (1941); for some substituted diphenyls in which the phenyl groups are held close to a planar configuration, he finds that the strength of the absorption is about equal to that of diphenyl itself.

Most of the visible and near-ultraviolet spectra of conjugated organic molecules undoubtedly belong to the V, N class. In such cases it may be expected that LCAO computations like those discussed here will give results corresponding to observation, provided the molecular model is correctly chosen. Where this is uncertain, such computations, along with measured intensities, should help to decide it.

* Here *cis* and *trans* refer to the arrangement about the central double bond, and the two forms can be obtained as separate isomers.

(g) *Empirical correction factors*

In comparing computed and observed intensities in table 5 we must keep in mind three difficulties:—(i) the computations are all by the LCAO MO method, and we may therefore expect to find that they differ systematically from the observed; (ii) in making the comparison we assume that the upper state of the observed transition is indeed the V state to which the computations refer; (iii) the correct nuclear configuration is not always certain. In view of these considerations, the best molecules for a reliable determination of the ratios between observed and computed f values, together with the values of these factors, are:

	Molecule	Correction factor
Chain	1,3-butadiene	0·40
	2,3-dimethyl-1,3-butadiene	0·45
Ring	Cyclopentadiene	0·30
	Cyclopentadiene (vapour)	0·32
	Cyclohexadiene	0·30
	Cyclohexadiene (vapour)	0·29
	Styrene	0·30
	Diphenyl	0·30
Ring-chain	Stilbene (*trans*)	0·34
	Diphenyl butadiene	0·30
	Diphenyl octatetraene	0·27

All observed f values used for the correction factors were measured in solution, except as otherwise indicated. The computed f values were obtained from the computed D and the observed ν_{max}, the latter being first corrected to vapour, if measured in solution. From the constancy of the correction factor it appears that we have correctly identified the electronic states involved. The small decrease with length for the diphenyl polyenes probably indicates that the LCAO method leads to systematically greater errors in large molecules (Mulliken, 1939 g), or possibly it may indicate that there is an increasing mixture of *cis* forms for the larger molecules. For very long chains, if these have the *trans* form, as seems likely, the correction factor falls considerably; thus for β-carotene, assumed *trans*, the factor is 0·15.

It would, of course, be desirable to have more observational data for the comparison—in particular for chain molecules. The two values found above for this factor are 0·40 and 0·45; since both end-substituted butadienes indicate (though uncertainly) a factor smaller than 0·4, we have adopted 0·4 as the value of the factor for chain molecules of moderate size. For rings the factor is 0·3. The computed intensities in table 5 A have accordingly been corrected, before comparison with experiment, by the factor 0·4, except for the cyclic dienes. The latter, and all intensities in table 5 B, have been corrected by 0·3. These factors may be compared with the factor for O_3 (gas)—0·34 from table 2.

§8. MOLECULAR REFRACTIVITIES AND OTHER OPTICAL PROPERTIES

(a) *Molecular refractivities*

In addition to the directly observed f values of the methyl butadienes and cyclic dienes, there exist indirect measurements in the form of molecular refractivities. According to classical dispersion theory, the molecular refractivity M for any substance (liquid or vapour), at a frequency ν where the substance is transparent, is given by

$$M=(Ae^2/3\pi mc^2)\Sigma_j f_j/(\nu_j^2-\nu^2)=1{\cdot}80\times10^{10}\Sigma_j f_j/(\nu_j^2-\nu^2), \qquad \ldots\ldots(37)$$

where A equals Avogadro's number and f_j is the oscillator strength of the absorption region whose maximum of absorption is at ν_j, infra-red as well as ultra-violet bands being included.

Existing empirical data on the refractivities of hydrocarbons for sodium D light can be expressed by (Mulliken, 1939f)

$$M=1{\cdot}705N_{CH}+1{\cdot}209N_{CC\sigma}+2{\cdot}942N_{CC\pi}+E, \qquad \ldots\ldots(38)$$

where N_{CH}, $N_{CC\sigma}$, and $N_{CC\pi}$ are the respective numbers of C – H bonds, C – Cσ bonds, and C – C unsaturation-electron (π) bonds, and E is what is left over, the " exaltation ". The coefficients in equation (38) have been adjusted so that E averages zero for normal unconjugated molecules. For conjugated molecules, E often has positive values of considerable size (cf. table 6), and it is generally accepted that this is a result mainly of the conjugation (Hückel, 1935). A formulation of bond refractivities, essentially the same as equation (38), has also been given by Denbigh (1940); it leads to exaltations differing inappreciably from those derived using equation (38). Lowry and Allsopp (1937, and earlier papers) have studied the relation of E in conjugated molecules to the ultra-violet absorption spectrum, and they and Hückel (1935) have discussed the relation of E and the spectrum to the nuclear configuration of the molecule. As one of the writers has shown (Mulliken, 1939f), if E arises mainly from conjugation, it should be possible to account for observed E values largely by means of the terms in equation (37), which correspond to V, N transitions of unsaturation electrons. We give here a revised discussion, including the results of new computations.

If we attribute E exclusively to the unsaturation electrons, we can obtain " observed " values of the partial molecular refractivity ΔM arising from the latter (cf. table 6, column four). These should be given theoretically by

$$\Delta M=\Delta M_{V,N}+\Delta M_{k,N}= 1{\cdot}80\times10^{10}[\Sigma_j f_{V_j,N}/(\nu^2_{V_j,N}-\nu^2)+\Sigma_k f_{k,N}/(\nu^2_{k,N}-\nu^2)] \qquad \ldots.(39)$$

In equation (39), ΔM is expressed as a sum of two parts corresponding respectively to V, N and to other (mostly high-frequency) transitions of the

unsaturation electrons. Although the second of these parts includes contributions from infra-red frequencies, these are so small for visible ν's that they may be neglected.

Table 6. Molecular refractivities $(M)^a$

Molecule	Model	Observed[b]			Computed[c]	
		Total M	ΔM	E	$\Delta M_{V,N}$	E
Propylene	—	—	—	—	1·8	0
1,3-butadiene	*trans*	21·1 ±0·1	7·3	1·4	5·0	1·4
2,3-dimethyl-1,3-butadiene	*trans*	29·72±0·05	6·63	0·75	4·1	0·5
2-methyl-1,3-butadiene	*trans*	25·41±0·05	6·93	1·04	5·2	1·6
	cis	—	—	—	3·7	0·1
1-methyl-1,3-butadiene	*trans*	25·8 ±0·1	7·3	1·4	5·6	1·9
1,4-dimethyl-1,3-butadiene	*trans*	30·81±0·05	7·72	1·84	6·3	2·6
1,3-cyclopentadiene	*cis*	21·85±0·05	5·57	−0·31	2·9	0·8
1,3-cyclohexadiene	*cis*	26·74±0·05	5·85	−0·03	3·6	0·0
Benzene	—	26·14±0·005	8·65	−0·17	4·7	0·8

(*a*) All data and computations refer to refractivities for sodium *D* light.

(*b*) From Mulliken (1939 f). ΔM obtained by subtracting $1{\cdot}705\ N_{CH} + 1{\cdot}209\ N_{CC\sigma}$ from the total M, E by subtracting $2{\cdot}942\ N_{CC\pi}$ from ΔM. Cf. equation (38).

(*c*) E (computed) $= \Delta M_{V,N} - \Delta M^{\circ}_{V,N}$; see equation (40) and text. $\Delta M^{\circ}_{V,N}$ is taken as $\Delta M_{V,N}$ for propylene times the number of double bonds (two, for the dienes).

The exaltation E corresponds to an increased ΔM for conjugated as compared with unconjugated molecules. Presumably most of this increase comes from the first summation in equation (39). For the moment let us neglect any change in the second part. We then need only to compute $\Delta M_{V,N}$ for the actual conjugated molecule, and to subtract from this the corresponding quantity $\Delta M^{\circ}_{V,N}$ computed for the same molecule in a hypothetical unconjugated state:

$$E = \Delta M_{V,N} - \Delta M^{\circ}_{V,N} \qquad \ldots\ldots (40)$$

$\Delta M^{\circ}_{V,N}$ can be obtained by multiplying the total number of double bonds in question by the value of $\Delta M_{V,N}$ computed for an unconjugated molecule containing one double bond. For this it would be natural to use ethylene, but since equation (38) is based on taking $E = 0$ for a *typical* unconjugated molecule, which usually contains substituent groups, we have chosen propylene as more nearly representative. The results are given in table 6; our computed ΔM's are based on the observed ν and the LCAO—computed $f_{V,N}$'s after correcting the latter by suitable empirical factors (cf. §7(g)). These corrected f values are mostly from table 5. In cases where ν values are not known from observation, as for most V, N transitions at higher frequencies than V_1, N, they have been estimated. Unfortunately, these estimates are very uncertain; but since V_2, N and V_3, N are strong only in the *cis*-dienes, this is important only for them.

Considering the various simplifying assumptions and other uncertainties, the agreements between computed and observed exaltations in table 6 are very

satisfactory, both as to absolute values and as to mode of variation with molecular structure. Although the molecules of non-cyclic dienes probably exist predominantly in the *trans* form, space models suggest that for isoprene, the *cis* form has a relatively high stability. We give in table 6 computations for both *cis*- and *trans*-isoprene; the computed and observed exaltations would agree if we assume that the observations refer to a mixture of 60% *trans*, 40% *cis*—the same proportions are necessary to give agreement between computed and observed intensity for the V_1, N transition (table 5). In all the substituted butadienes, and in propylene, our computed $\Delta M_{V,N}$'s and E's allow for hyperconjugation, which is seen to be of considerable importance in determining the value of E.

Although it is seen that the observed exaltations E can be explained rather well by the $\Delta M_{V,N}$'s, it is to be expected theoretically that contributions to E come also from other sources. In particular, the $\Delta M_{k,N}$'s should make small contributions, and the C—H and C—Cσ electrons also may contribute appreciably, sometimes to increase, sometimes to decrease E. This is seen experimentally from the fact that E sometimes differs considerably from zero even when no double bonds at all are present. It is understandable theoretically in terms of hyperconjugation. In general, however, the V, N and other transitions of the σ electrons lie at higher frequencies, and are less affected by changes in the molecule, than is the case for the unsaturation electrons.

It is of interest that, although $\Delta M_{V,N}$ amounts to half or two-thirds of ΔM in the various molecules listed in table 6, the corresponding $\Sigma f_{V,N}$ is only about one-sixth of the total Σf for the unsaturation electrons.* From the $\Delta M_{k,N}$ part of ΔM, together with the $\Sigma f_{k,N}$ found by subtracting $\Sigma f_{V,N}$ from the Σf, we can determine a mean effective frequency in equation (39) for the k, N transitions; this comes out with values of about 170,000 cm^{-1}, much beyond the average ionization limit for unsaturation electrons (about 80,000 cm^{-1}).

Broadly speaking, large positive E values arise from conjugation and hyperconjugation in molecules of *trans* form, while *cis* or ring forms give E values near zero. And this is seen to be primarily a result of the spread in frequency among the different V, N transitions, together with the large concentration of intensity in their longest wave-length member V_1, N for the *trans* model.

(b) *Polarizability and optical anisotropy*

The partial theoretical explanation of molecular refractivities just given, could be paralleled by similar discussion of other related properties. The interesting relations between refractivity, bond length, bond order, and anisotropy discussed by Denbigh (1940) are beyond the scope of this article. In the case of hydrogen gas, dispersion, polarizability, and optical anisotropy have already been discussed in § 5.

* If the unsaturation electrons were independent of the other electrons, the usual f sum rule would make their Σf equal to their number. Actually their Σf must be greater because of virtual absorption transitions with negative frequency, going to 2*s* and 1*s* AO's.

In polyene chain molecules of *trans* form, the low frequency and high f value of the V_1, N transition should lead to an abnormally high polarizability (cf. equation (28)). Moreover, this should be highly anisotropic, with an extremely large maximum value lying approximately along the longest direction of the molecule, as can be seen from vector equations analogous to equations (36) for butadiene. This high longitudinal polarity may be connected with the tendency of such molecules to polymerize (Mulliken, 1939.g) and should also have interesting consequences for the Raman effect (Mulliken, 1939 h).

In conjugated molecules of more compact shape, including benzene and other aromatic molecules, the V, N unsaturation electron transitions always give rise to high polarizability *in the plane* of the molecule, even though not in one special direction. This high polarizability has been recognized for some time, although not specifically attributed to V, N transitions. Briegleb (1936) believes that it is the cause of the formation of many loose molecular compounds between aromatic molecules. Pauling (1939) has expressed similar views. Applications to crystal optics and other subjects have been made by various authors, but we shall not attempt to discuss these.

Long polyene chain molecules should be highly anisotropic optically. Consider, for example, a crystal built up of elongated conjugated polyene molecules aligned more or less parallel. The absorption coefficient in the $N \rightarrow V_1$ region, and the refractive index at longer wave-lengths, should be much higher for light with electric vector vibrating parallel to the long molecule-axis than for light polarized in either of the directions perpendicular to this. There is evidence that crystals of the carotene-like substance methylbixin do indeed show exactly this behaviour (Mulliken, 1939 h). Further, the refractive index is least for polarization perpendicular to the plane of the molecule, as would be expected, since *all* $N \rightarrow V$ transitions should be polarized in the plane of the molecule.

§9. VARIATION OF TRANSITION MOMENT WITH INTERNUCLEAR DISTANCES: FORBIDDEN TRANSITIONS

(a) *Variation of Q and D with distance*

Up to this point the quantities Q and D of equation (1) have been computed for molecules assumed fixed in their equilibrium configurations. Actually, since the forms of the electronic wave functions vary somewhat with the internuclear distances, so does the dipole moment eQ_{kl} of a molecular transition, as well as the ordinary dipole moment eQ_{ll} of the normal state.* Thus for any Q_{kl},

$$Q = Q_0 + \Sigma_n Q_1^n q_n + {}_m\Sigma\Sigma_n Q_2^{mn} q_m q_n + \ldots, \qquad \ldots\ldots(41)$$

* The absolute intensities of infra-red bands are largely determined by the Q's in equation (41) for state N(Q_{ll}'s): the pure rotation spectrum intensities are proportional to Q_0^2, and the intensity of the fundamental band associated with the nth normal mode of vibration is proportional to $(Q_1^n)^2$. Thus equation (41) for state N is of great interest in connection with infra-red spectra and dispersion. We shall, however, confine outselves here to electronic spectra.

where the q's are a set of co-ordinates describing displacements of the nuclei from equilibrium—for example vibrational normal co-ordinates. In diatomic molecules, there is only one q, namely $\sqrt{\mu}(r-r_e)$.

Our discussion has hitherto assumed that the total intensity of an electronic transition is proportional to Q_0^2 and independent of the amount of vibration and rotation. The variation of Q with the co-ordinates is neglected also in the ordinary Franck-Condon procedure for determining the relative intensities of different vibrational bands in an electronic transition. The effect of the Q_i's with i greater than zero should, however, sometimes be important, particularly if Q_0 is small.

For a few diatomic molecules, the variation of Q with r has been computed. The first such computation was made by James (1934, 35), and James and Coolidge (1933, 39) in connection with their study of the intensity distribution in the continuous hydrogen *emission* spectrum arising from the transition

$$(\sigma_g 1s)(2s\sigma_g),\ ^3\Sigma^+_g \rightarrow (\sigma_g 1s)(\sigma_u 1s),\ ^3\Sigma^+_u.$$

Their computations far exceed in accuracy all others discussed in this *Report*, having been based on extremely accurate wave functions (much better than is possible for any AO or MO approximation). The computed D values are as follows:

r (A.) =	0·688	0·794	0·899	0·989	1·058	1·111	1·217	1·323	1·429	1·534
D (A^2) =	0·827	0·632	0·465	0·356	0·287	0·266	0·200	0·154	0·127	0·105

With such a rapid variation of D with r, the transition probabilities are appreciably dependent on the initial state of vibration. Accordingly, James and Coolidge have computed them for several vibrational levels of the initial (here upper) state. They find for $v=0$, 1, 2, 3, respectively, mean lifetimes of 1·19, 1·10, 1·01, $0{\cdot}97\times10^{-8}$ sec. (v = vibrational quantum number; the mean lifetime for emission is inversely proportional to the mean effective D values times ν^4). The variation with v is only moderate in spite of the rapid variation of D with r. For other molecules than hydrogen, since they in general have smaller vibrational amplitudes, the effect should be even smaller. This gives assurance that the procedure so far followed in this *Report* of using D values computed for equilibrium values of r should seldom, if ever, lead to large errors.

James and Coolidge also investigated the validity of the simple Franck-Condon principle for the same transition. They find the intensity distribution within the spectrum to be surprisingly little affected by the variation of D with r. In particular, the position of ν_{max} of the continuum is found to be shifted by only about 2000 cm^{-1} for the variable as compared with a constant D. One may expect that the effect will be smaller than this in the spectra of most molecules, even with extreme variation of D with r.

Computations have also been made on the Q, N transitions in the halogen molecules (Mulliken, 1940 b; Sayvetz, unpublished). The results by both AO and LCAO approximations are summarized in table 7. They show a variation

with r much more rapid even than that just discussed for hydrogen. This is connected with the fact that as $r \to \infty$, the transition goes over into one forbidden for the separate atoms. Such a relation is very frequent among valence-shell transitions in molecules.

In spite of the rapid variation of D with r, the variation in D with vibrational quantum number should not be very great. Further, ν_{max} of the Q, N absorption continuum in Cl_2 is computed by Sayvetz (unpublished) to have been shifted only about $+300$ cm.$^{-1}$ as compared with the location it would have had for a constant D.

In these computations r enters mainly through a parameter γ equal to rZ/a_0n^*. Therefore if D is sensitive to r, it is equally sensitive to the assumed Z of the AO's used. We have made use of this in § 6 in concluding that a small change in the originally assumed Z's would bring about good agreement between observed and computed D's in F_2, Cl_2, and Br_2.

Table 7. Computed D as a function of internuclear distance r

Molecule	r (A.)[a]	Dipole strength (A.^{2}) LCAO	AO	Observed
	1·20	0·0216	0·0040	
	1·28	0·0138	0·0022	
	1·40	0·0072	0·00080	
F_2	**1.435**	0·0060	0·00058	0·00068
	1·50	0·0040	0·00028	
	1·60	0·0032	0·000092	
	1·90	0·0866	0·0208	
Cl_2	**1.983**	0·0634	0·0138	0·0056
	2·10	0·0381	0·0075	
	2.30	0·0182	0·0016	

(*a*) The equilibrum distance r_0 is indicated by bold-faced type.

(b) *Forbidden electronic transitions in polyatomic molecules*

If $Q_0 = 0$ in equation (41), then, so far as electric dipole transitions are concerned, the transition is forbidden by the electronic selection rules. These are determined by the symmetry of the equilibrium configuration. In diatomic molecules, where vibration cannot change the symmetry, the entire Q of equation (41) is then zero. In polyatomic molecules, on the other hand, there are always modes of vibration which destroy the equilibrium symmetry; accordingly, not all the Q_i's are zero when $Q_0 = 0$. As a result, such " forbidden " transitions actually occur with appreciable, although usually low, intensity (Herzberg and Teller, 1933).

To determine the intensities, one must make use of the integrals

$$P = \int \Omega_l Q \Omega_k d\tau, \qquad \ldots\ldots(42)$$

where Ω_l and Ω_k are vibrational wave functions of the initial and final states. Then, for a given initial state Ω_l, we obtain D for the electronic transition by

summing the quantities P^2 over all possible final vibrational states, and multiplying by G if the electronic states require it.

Most of the intensity of such a forbidden transition comes from transitions in which Ω_k and Ω_l differ by just one quantum of one or another of the unsymmetrical modes of vibration. The total strength for any given initial state, in which the nth normal mode of vibration (frequency ν_n) is excited to quantum number v_n, is found to be given by

$$D=(Gh/8\pi^2c)\Sigma_n(2v_n+1)(Q_1^n)^2/\nu_n, \qquad \ldots\ldots(43)$$

which for forbidden transitions takes the place of equation (1).* Since with $Q_1^n=\partial Q/\partial q_n$ and q_n is of the dimensions of $\sqrt{m}l$ (m = nuclear mass, l = some internuclear distance), $(Q_1^n)^2$ has the dimensions of $1/m$. Its order of magnitude may go as high as $1/m$, since $\partial Q/\partial l$ may be as large as unity (compare, for example, the values of Q^2 as a function of distance for the hydrogen continuum in § 9 (a)). Equation (43) then would give a D of approximately 0·01 Å^2 for a vibrationless initial state (all V_n's zero), but usually much smaller values are to be expected. Equation (43) predicts that, in contrast to D for allowed transitions, D here increases with increasing vibration in the initial state, hence increases with temperature.

The intensities of forbidden transitions are, in a sense, stolen from neighbouring allowed transitions. In principle, the coefficients Q_1^n can be obtained by perturbation theory if all the electronic wave functions and energies of a molecule are known. If for equilibrium the upper state electronic wave function is Ψ_0, the manner in which this changes with a nuclear co-ordinate q_n may be expressed in terms of mixing of other electronic wave functions into Ψ_0 (under the influence of the perturbing potential, proportional to q_n, which is developed when q_n departs from zero) (Herzberg and Teller, 1933). Since the extent of

* Let us consider the simple case where just one unsymmetrical vibration mode, with normal co-ordinate q_n in the initial state, is responsible for the occurrence of the bands. The effective Q is then

$$Q=q_n(Q_1^n+\Sigma_m Q_1^m q_m+\ldots).$$

To the same degree of approximation that we used in putting $Q=Q_0$ for allowed transitions, here we may use $Q=q_n\,Q_1^n$. We then evaluate P in equation (42) for the simplest case of an initial state with all vibrational quantum numbers zero. We first note that the integral contains a factor $\Psi_0''q_n$, where Ψ_0^n is the vibrational wave function, with quantum number v and frequency $c\nu_n$, associated with the normal mode of vibration q_n. From the properties of the well-known harmonic-oscillator wave functions, $\Psi_v''q_n$ may be written as $(h/8\pi^2c\nu_n)^{1/2}\Psi_1''$. Then integrating, squaring, and summing over all final vibrational states, we get the D due to q_n. The summation over the final vibrational states gives

$$(h/8\pi^2c\nu_n)(Q_1^n)^2\Sigma_k\left[\int\Psi_1''\Psi_k''dq_n\int\Omega_k'\Omega_l'd\tau'\right]^2.$$

where Ω' is the part of Ω of equation (42) which remains after factoring out Ψ''. It can be shown, after expanding the final state Ψ's and Ω's in terms of those of the initial state, that the summation equals unity. This is immediately obvious in the special case where the initial and final equilibrium nuclear configurations are identical; here only one vibrational transition, with $\Omega_k'\equiv\Omega_l$ and $\Psi_k''\equiv\Psi_1''$ occurs. Finally, a summation over all q_n's yields equation (43) for the case that all v_n's are zero. The derivation of the general case proceeds in a similar way.

such mixing is inversely proportional to the distance apart of the perturbed and perturbing electronic levels, Q_1^n varies inversely as this distance, and $(Q_1^n)^2$ inversely as its square. If one of the nearby perturbing electronic levels is one which receives an intense allowed transition from the ground state, this transition will be predominantly responsible for Q_1^n, and it may then be said that D of the forbidden transition is stolen largely from this allowed transition.

Some of the best known spectra of organic compounds are probably forbidden transitions. This is true * of the weak absorption near λ 2900 (f about 0·0003) characteristic of aldehydes and ketones (Mulliken, 1935 b; McMurry and Mulliken, 1940; McMurry, 1941 a, 41 b). It is probable that the observed strength of this transition is drawn mainly from the allowed Q, N transition near λ 1900 (cf. § 6 (*c*)).

Another well-known forbidden transition is the benzene absorption system near λ 2600 (f = 0·0016). Here the identification as a forbidden transition has been established beyond any question (Sponer, Nordheim, Sklar and Teller, 1939; Sponer and Teller, 1941). There can also be little doubt that the λ 2600 transition steals its strength from the very intense V, N transition (cf. table 5) near λ 1800. Another transition in benzene which is probably a forbidden one (Sponer and Teller, 1941), in spite of its high f value (0·13), occurs near λ 1950. Taking into account the inverse square law mentioned above, and also an additional effect which enters (Herzberg and Teller, 1933) when approach is very close, the close proximity of the λ 1950 system to the V, N region may suffice to account for its high intensity.

The λ 2700 spectra of the substituted benzenes discussed in § 6 (d) and table 4 are examples of transitions where forbiddenness in a parent molecule has been removed by loss of equilibrium symmetry as a result of substitution. In the weakest of these spectra, however, certain features of the structure show (Sponer and Wollmann, 1940; Sponer and Teller, 1941) that an appreciable part of the observed strength is still of the forbidden Q_1^n type. This is an illustration of the fact that the Q_1 terms in equation (41) can become important when Q_0 is unusually small.

REFERENCES

Bayliss, N. S. and Rees, A. L. G., 1939. *Trans. Faraday Soc.* **35**, 792. (Contains also references to other works on bromine absorption.)

Booker, H., Evans, L. K. and Gillam, A. E., 1940. *J. Chem. Soc.* 1453.

Briegleb, G., 1936. *Z. phys. Chem.* B, **31**, 58.

Chako, N. Q., 1934. *J. Chem. Phys.* **2**, 645.

Child, C. L. and Walker, O. J., 1938. *Trans. Faraday Soc.* **34**, 1506.

Coulson, C. A., 1937. *Trans. Faraday Soc.* **33**, 1484.

Denbigh, K. G., 1940. *Trans. Faraday Soc.* **36**, 936.

Easthope, C. E., 1936. *Proc. Camb. Phil. Soc.* **32**, 260.

* Strictly speaking, this transition is not quite forbidden except in the most symmetrical of these molecules, but apparently the "allowed" strength resulting from asymmetry is negligible compared with the "forbidden" strength (cf. final paragraph of this section).

FÖRSTER, T., 1939. *Z. Elektrochem.* **45**, 548. [Review.]
HAUSSER, K. W., KUHN, R., SMAKULA, A. and HOFFER, M., 1935. *Z. phys. Chem.* B, **29**, 364.
HENRI, V. and PICKETT, L. W., 1939. *J. Chem. Phys.* **7**, 439.
HERZBERG, G. and TELLER, E., 1933. *Z. phys. Chem.* B, **21**, 410.
HIRSCHFELDER, J. O., 1935. *J. Chem. Phys.* **3**, 555.
HÜCKEL, E., 1937. *Z. Elektrochem.* **43**, 752, 827. [Review.]
HÜCKEL, W., 1935. *Theoretische Grundlagen der organischen chemie*, Akad. Verlag M. B. H. Leipzig, second edition, pp. 83–106 ; 139–158 tables.
JAMES, H. M., 1934. *J. Chem. Phys.* **2**, 794.
JAMES, H. M., 1935. *J. Chem. Phys.* **3**, 9.
JAMES, H. M. and COOLIDGE, A. S., 1933. *J. Chem. Phys.* **1**, 825.
JAMES, H. M. and COOLIDGE, A., 1939. *Phys. Rev.* **55**, 184.
JONES, R. N., 1941. *J. Amer. Chem. Soc.* **63**, 1658.
KOHLRAUSCH, K. W. F. and STOCKMAIR, W., 1935. *Z. phys. Chem.* B, **29**, 292.
LENNARD-JONES, J. E., 1937. *Proc. Roy. Soc.* A, **158**, 280.
LEWIS, G. N., MAGEL, T. T. and LIPKIN, D., 1940. *J. Amer. Chem. Soc.* **62**, 2973.
LOWRY, T. M. and ALLSOPP, C. B., 1937. *Proc. Roy. Soc.* A, **163**, 356.
LYDDANE, R. H., ROGERS, F. T. Jr. and ROACH, F. E., 1941. *Phys. Rev.* **60**, 281.
MCMURRY, H. L., 1941 a. *J. Chem. Phys.* **9**, 231.
MCMURRY, H. L., 1941 b. *J. Chem. Phys.* **9**, 242.
MCMURRY, H. L. and MULLIKEN, R. S., 1940. *Proc. Nat. Acad. Sci., Wash.*, **26**, 312.
MROWKA, B., 1932. *Z. Phys.* **76**, 300.
MROWKA, B., 1933. *Z. Phys.* **84**, 448.
MROWKA, B., 1939. *Eucken-Wolf's Hand- und Jahrbuch der Chem. Physik*, **10** (iii), 25.
MULLIKEN, R. S., 1935 a. *J. Chem. Phys.* **3**, 375.
MULLIKEN, R. S., 1935 b. *J. Chem. Phys.* **3**, 564.
MULLIKEN, R. S., 1939 a. *J. Chem. Phys.* **7**, 14.
MULLIKEN, R. S., 1939 b. *J. Chem. Phys.* **7**, 20.
MULLIKEN, R. S., 1939 c. *J. Chem. Phys.* **7**, 121.
MULLIKEN, R. S., 1939 d. *J. Chem. Phys.* **7**, 339.
MULLIKEN, R. S., 1939 e. *J. Chem. Phys.* **7**, 353.
MULLIKEN, R. S., 1939 f. *J. Chem. Phys.* **7**, 356.
MULLIKEN, R. S., 1939 g. *J. Chem. Phys.* **7**, 364.
MULLIKEN, R. S., 1939 h. *J. Chem. Phys.* **7**, 570.
MULLIKEN, R. S., 1940 a. *Phys. Rev.* **57**, 500.
MULLIKEN, R. S., 1940 b. *J. Chem. Phys.* **8**, 234.
MULLIKEN, R. S., 1940 c. *J. Chem. Phys.* **8**, 382.
MULLIKEN, R. S. and RIEKE, C. A., 1941. *J. Amer. Chem. Soc.* **63**, 1770.
MULLIKEN, R. S., RIEKE, C. A. and BROWN, W. G., 1941. *J. Amer. Chem. Soc.* **63**, 41.
OLDENBERG, O. and RIEKE, F. F., 1938. *J. Chem. Phys.* **6**, 443–444.
O'SHAUGHNESSY, M. T. and RODEBUSH, W. H., 1940. *J. Amer. Chem. Soc.* **62**, 2906.
PAULING, L., 1939. *Proc. Nat. Acad. Sci., Wash.*, **25**, 577.
PAULING, L. and WILSON, E. B., Jr., 1935. *Introduction to Quantum Mechanics* (New York : McGraw-Hill) ; earlier references on p. 353.
PICKETT, L. W., PADDOCK, E. and SACKTER, E., 1941. *J. Amer. Chem. Soc.* **63**, 1073.
PRICE, W. C. and TUTTE, W. T., 1940. *Proc. Roy. Soc.* A, **174**, 207.
PRICE, W. C. and WALSH, A. D., 1940. *Proc. Roy. Soc.* A, **174**, 220.
RABINOWITCH, E. and WOOD, W. C., 1936. *Trans. Faraday Soc.* **32**, 540.
SCHEIBE, G. and FRÖMEL, W., 1937. *Eucken-Wolf's Hand- und Jahrbuch der Chem. Physik*, **9** (iv), 144.
SCHEIBE, G. and GRIENEISEN, H., 1934. *Z. phys. Chem.* B, **25**, 54.
SCHOMAKER, V. and PAULING, L., 1939. *J. Amer. Chem. Soc.* **61**, 1769.

SKLAR, A. L., 1939. *J. Chem. Phys.* **7**, 984.
SMAKULA, A., 1934. *Z. angew. Chem.* **47**, 657.
SMAKULA, A. and WASSERMANN, A., 1931. *Z. phys. Chem.* A, **155**, 353.
SPONER, H., NORDHEIM, G., SKLAR, A. L. and TELLER, E., 1939. *J. Chem. Phys.* **7**, 207.
SPONER, H. and TELLER, E., 1941. *Rev. Mod. Phys.* **13**, 75.
SPONER, H. and WOLLMAN, S. H., 1940. *Phys. Rev.* **57**, 1078 A.
STARK *et al.*, 1913. *Jahrbuch der Radioaktivität*, **10**, 170.
STEENSHOLT, G., 1935. *Z. Phys.* **93**, 620.
STUART, H. A., 1936. *Eucken-Wolf's Hand- und Jahrbuch der Chem. Physik*, **8** (ii), 32, **48**.
SUGIURA, Y., 1927. *J. de Physique*, **8**, 117.
WHEELER, J. A., 1933. *Phys. Rev.* **43**, 258.
WOLF, K. L. and HERZFELD, K. F., 1928. *Handbuch der Physik*, **20**, 492.

SYMMETRY PROPERTIES OF NUCLEAR LEVELS

By EUGENE P. WIGNER,
Palmer Physical Laboratory, Princeton, New Jersey,
AND
EUGENE FEENBERG,
Washington Square College, New York University, New York

§1. INTRODUCTION

THE theories of nuclear structure which form the subject of the present *Report* resemble in many respects the quantum theory of spectroscopy (Condon and Shortley, 1936). In one respect, the problem is much simpler: in the nuclear theory one would be quite satisfied with an adequate description of the normal state and of a few excited states, a calculation of their energies, spins and other properties. In atomic spectroscopy, on the other hand, one is also often interested in the properties of rather highly excited states. The main reason why the problem of nuclear structure is so much further from its solution than those problems encountered in spectroscopy is that hardly anything is known with certainty about the nature of nuclear forces. The forces between electrons, and electrons and nuclei, are, on the other hand, sufficiently well known for the purposes of spectroscopy.

Two kinds of consideration are used for the derivation of the most important relations of spectroscopy. The first kind (Wigner, 1931) can be based on an investigation of the symmetry properties of the system, or, expressing it more mathematically, on an investigation of the group of operators under which the Hamiltonian is invariant. This consideration leads to a classification of the energy levels, and to a description of their fine structure and other properties. It does not tell us, however, anything about the position of the levels. The second kind of consideration attempts, on the basis of a more specialized model, to predict the position of the energy levels. Thus, e.g., the first kind of consideration tells us that there are, among others, ^3D terms, and that these have a fine structure in which the distances from the middle term to the two other terms are in the ratio 2:3; it tells us how to calculate the Zeeman effect of these terms, and so on. The second kind of consideration permits us to predict that the lowest term of oxygen is a ^3D term, and that it is followed by a ^3S and a ^1P term (Slater, 1929).

The situation in the nuclear theories which will be discussed here is very similar to that sketched above for spectroscopy. There is only one essential

difference: one is able to predict the approximate position of the lowest level (although not all its characteristics) by an argument which uses the mathematical methods of the symmetry considerations. However, if one tries to derive a more complete level scheme, one is forced again to considerations very similar to those involving the configurations and the Hartree model of ordinary spectroscopy.

§ 2. QUANTUM NUMBERS OF NUCLEAR LEVELS

We shall begin with the first kind of consideration, which leads us to a classification of the nuclear energy levels by quantum numbers (Wigner, 1937 a; Hund, 1937 a). There are three degrees of detailedness with which such a classification can be carried out. The first one makes use only of the symmetry properties of space time and leads to the total (or inner) quantum number and the parity. The definition of these quantum numbers does not require any special assumptions concerning the nature of forces and is—just as in spectroscopy—quite rigorous.

The next step involves, both in spectroscopy and in nuclear theory, assumptions concerning the nature of forces. In ordinary spectroscopy, the assumption ordinarily made is that the energy is independent of the orientation of the spins. This leads to the definitions of a multiplet system S and azimuthal quantum number L. In nuclear theory, two distinct assumptions are possible. The first, less far-reaching, assumption postulates that the forces be the same between all combinations of nuclear particles (protons and neutrons). The second, further-reaching, assumption, which leads to the most detailed classification of nuclear energy levels, is the same as that of spectroscopy: that the forces are independent of the orientation of the spin. This last approximation, in which the potential energy is independent both of the orientation of the spin and the nature of the particle (i.e. whether it is a proton or a neutron), will be called *1st approximation*. The somewhat less drastic assumption, which allows for a dependence of the energy on the orientation of the spin (but still assumes that it is independent of the nature of the particle), will be called *2nd approximation*. The Hamiltonian, which gives the same interaction between proton-neutron, proton-proton and neutron-neutron pairs, is often called a symmetric Hamiltonian.

Neither of these assumptions is, of course, strictly correct. The assumption that the nuclear forces between two protons are the same as between two neutrons is due to Heisenberg (1932 a, 32 b). That they are also equal to those acting between a proton and a neutron has been suggested by the experimental result that the potential determining the scattering of neutrons and of protons by protons appears to be the same (Breit, Condon and Present, 1936; Breit and Stehn, 1937; Breit, Thaxton and Eisenbud, 1939: these papers contain quite comprehensive references to the experimental works). The above statements refer to the equality of nuclear forces alone: between a pair of protons the electrostatic force acts in addition to these (and plays, indeed, a major rôle also in the

above scattering experiments). Thus the existence of the electrostatic (or " Coulomb ") energy gives a natural limitation to the validity of the assumptions used in the 2nd approximation. It will be assumed in the present *Report*—although the experimental evidence for this is somewhat indirect—that the electrostatic forces are the only important forces by which the interactions between the different nuclear particles differ (Breit and Feenberg, 1936).

The assumption that the forces do not depend greatly on the orientation of the spin appears, at first sight, quite reasonable. However, it leads to some

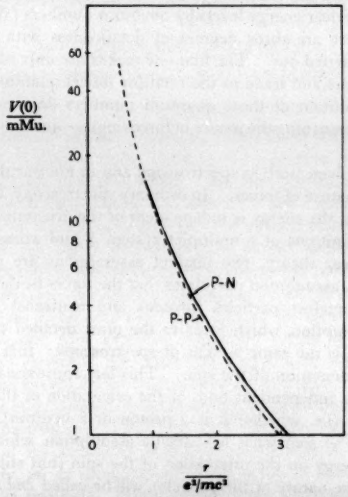


Figure 1. Connection between width and depth of potential well, as obtained from scattering experiments (cf. references: Breit, Condon and Present, 1936; Breit and Stehn, 1937; Breit, Thaxton and Eisenbud, 1939). P-P for proton-proton, P-N for proton-neutron interaction.

consequences which seem to be contradicted by experiment. It has, among others, the consequence that the azimuthal quantum number L and the multiplicity S should be good quantum numbers. It would then follow that the charge distribution in a ^3S state is spherically symmetric. The normal state of the deuteron is supposed to be a ^3S state and to show, therefore, no quadrupole moment. However, a quadrupole moment of considerable size has been found by Rabi and others (Kellogg, Rabi, Ramsey and Zacharias, 1940*).

* Cf. also Nordsieck (1940). The effect was found, qualitatively, even before, by Farkas and Sandler (1939); cf. also Casimir (1940).

A more detailed consideration on the basis of the 1st approximation also shows that the elements ^{2}H, ^{6}Li, ^{10}B, ^{14}N should have the same magnetic moment—which is again not borne out by experiment (see Millman and Kusch, 1941). It is evident, therefore, that the 1st approximation is at least a very inaccurate one and must always be supplemented by a consideration of perturbations caused by spin-dependent forces.

The quadrupole moment of ^{2}H has suggested that there are, perhaps, no spin-independent forces at all (Bethe, 1939, 40). It appears to us, however, that the weight of the experimental evidence is against this view (Rarita and Schwinger, 1941 a; Volkoff, 1942).

An enumeration of the interactions which do not involve the velocity of the particles to a higher power than the first has been given by Eisenbud and Wigner (1941).* They show that there are 18 types of interactions of this character. The 1st approximation would be rigorous if only two of these 18 kinds of interaction existed: ordinary forces and Majorana exchange forces (see below). There are eight types of interaction with which the 2nd approximation is compatible. The electrostatic force between the protons is, of course, not among these.

§3. ISOTOPIC SPIN

Heisenberg (1932 a) has given a very convenient formalism for dealing with the two kinds of particles which are present in a nucleus. He introduced, for this purpose, the isotopic spin variable. We shall denote this variable by τ. The value $\tau=1$ represents a neutron and $\tau=-1$ represents a proton in exactly the same way that $\sigma=1$ represents a particle with a spin in the direction of the z axis and $\sigma=-1$ a particle with a spin in the opposite direction.

The total wave function of a system with n particles contains, in addition to the space co-ordinates of these particles, two spin co-ordinates for each

$$\psi(\mathbf{r}_1, \sigma_1, \tau_1, \mathbf{r}_2, \sigma_2, \tau_2, \ldots, \mathbf{r}_n, \sigma_n, \tau_n). \qquad \ldots\ldots(1)$$

The Pauli principle makes it possible to assume that this wave function is antisymmetric (Bartlett, 1936; Elsasser, 1936) so that, e.g.,

$$\psi(\mathbf{r}_2, \sigma_2, \tau_2, \mathbf{r}_1, \sigma_1, \tau_1, \mathbf{r}_3, \sigma_3, \tau_3, \ldots) = -\psi(\mathbf{r}_1, \sigma_1, \tau_1, \mathbf{r}_2, \sigma_2, \tau_2, \mathbf{r}_3, \sigma_3, \tau_3, \ldots) \qquad \ldots\ldots(1a)$$

The wave function (1) will describe a state of one single nucleus if it vanishes unless

$$\tfrac{1}{2}(\tau_1+\tau_2+\ldots+\tau_n)=T_\zeta \qquad \ldots\ldots(2)$$

has one definite value. This value is called the isotopic spin, and is half the difference between the number of neutrons and of protons in the nucleus

$$\tfrac{1}{2}(n_N-n_P)=T_\zeta. \qquad \ldots\ldots(2a)$$

Mathematically, the isotopic spin can be treated in exactly the same manner as the ordinary spin. One can define operators τ_ξ, τ_η, τ_ζ which have the same

* See also Cassen and Condon, 1936.

effect on the variable τ_1 as Pauli's spin operators s_{x1}, s_{y1}, s_{z1} have on σ_1. In particular, T_ζ is mathematically completely analogous to S_z, the z component of the spin angular momentum. This is, of course, a purely mathematical analogy, as is evident, for example, from the fact that states with a definite value of the x component of spin momentum are physically just as interesting as states with a definite value of the z component of this quantity. On the other hand, states with a definite value of

$$T_\xi = \tfrac{1}{2}(\tau_{\xi 1} + \tau_{\xi 2} + \ldots + \tau_{\xi n})$$

have no physical significance, while states with a definite value of the operator

$$T_\zeta = \tfrac{1}{2}(\tau_{\zeta 1} + \tau_{\zeta 2} + \ldots + \tau_{\zeta n})$$

are states of a definite nucleus.

§ 4. THE 2ND APPROXIMATION

In the 2nd approximation the Hamiltonian does not involve any of the τ operators. It therefore commutes with T_ξ, T_η and T_ζ. Hence, just as the quantum number J can be defined in ordinary spectroscopy on the basis of the commuting of the Hamiltonian with J_x, J_y, J_z, so a *total isotopic spin quantum number* T can be defined (Wigner, 1937 a) in our 2nd approximation. It is a common energy level for $2T+1$ states with T_ζ values ranging from $-T$ to T. This means that several isobaric nuclei have, in this approximation, terms with the same binding energy. In particular, if a nucleus with isotopic spin T_ζ has a term with a certain binding energy, all isobaric nuclei with a smaller absolute value of the isotopic spin will have a term with the same binding energy. We shall say that these terms form a T-multiplet.

The isotopic spins are integers for nuclei with even mass-number n and half integers for odd n. In addition to T, naturally, the total angular momentum J and the parity remain good quantum numbers and are the same for all $2T+1$ terms of a T-multiplet.

The above relations suggest the uniting of the energy-level diagrams of isobaric nuclei into a common figure. Figure 2*a* is a schematic diagram of this kind for an even n in which it is assumed that the 2nd approximation is strictly valid. The ordinates are negative binding energies, their zero is at some very high point. The dots, which have equal ordinates, correspond to energy levels of different nuclei which have, however, the same binding energies, J values and parities. They form a T-multiplet.

In later diagrams, the dots of the terms of a T-multiplet will be replaced by a line connecting them. Furthermore, the ordinates will represent masses rather than negative binding energies. Hence, because the mass of the neutron is larger than that of ^{1}H, the ordinates will increase with increasing T_ζ within the same T-multiplet. Figure 2*b* represents the same (imaginary) energy levels as are represented in figure 2*a*.

It was pointed out before that the greatest perturbation of the 2nd approximation can be attributed to the electrostatic energy. We shall see later that the

electrostatic energy of a nucleus is very nearly equal to that of a sphere in which n_P charges e are distributed with uniform probability over the sphere. The radius of the sphere is called the nuclear radius and will be denoted by R. We assume a constant particle density for all low-lying nuclear states—all these assumptions will be tested later on—which will make R proportional to $n^{1/3}$.

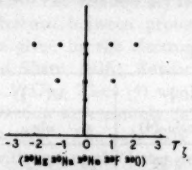


Figure 2*a*.

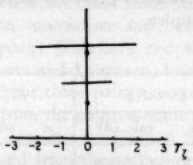


Figure 2*b*.

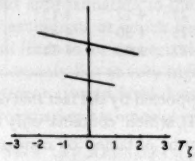


Figure 2*c*.

Figure 2. Common energy diagram for isobars (schematic). Figures 2*a* (for binding energy) and 2*b* (for masses) correspond to the 2nd approximation. Figure 2*c* takes the electrostatic interaction into account.

In this way we obtain for the masses of the different states of a T-multiplet

$$\left.\begin{aligned} M(T_\zeta) &= (M_N - M_H)T_\zeta + 3n_P(n_P - 1)e^2/5Rc^2 + \text{const.} \\ &= (M_N - M_H)T_\zeta + 3e^2(\tfrac{1}{2}n - T_\zeta)(\tfrac{1}{2}n - T_\zeta - 1)/5Rc^2 + \text{const.} \\ &= (M_N - M_H)T_\zeta - (3e^2/5Rc^2)T_\zeta(n - 1 - T_\zeta) + M(0). \end{aligned}\right\} \quad \ldots\ldots(3)$$

$M(T_\zeta)$ is the mass of the nucleus with isotopic spin T in the state of the T-multiplet in question. The first term on the right is caused by the masses of the neutron and of the ^{1}H atom, M_N and M_H, being different; the second term is due to the electrostatic energy. Expressing the masses in mMu. (1/16000th of the ^{16}O mass), (3) becomes (Wigner, 1937 b; Barkas, 1939)

$$M(T_\zeta) - M(0) = 0{\cdot}79T_\zeta - \tfrac{1}{4}1{\cdot}27T_\zeta(n - 1 - T_\zeta)/n^{1/3}. \quad \ldots\ldots(4)$$

The second numerical factor is adjusted in such a way that (4) gives close agreement with the observed mass differences of the isobaric pairs of table 1. Its value corresponds to a nuclear radius

$$R = 945e^2n^{1/3}/Mc^2 = 1{\cdot}46 \times 10^{-13}n^{1/3}\ \text{cm.}, \quad \ldots\ldots(5)$$

where M is 1/16th of the mass of ^{16}O. The electrostatic correction is taken into account in figure 2*c*, which is, otherwise, based on figures 2*a* and 2*b*.

Table 1* contains the mass differences of the normal states of isobaric pairs in which the number of protons and neutrons differs by one. ^{13}C–^{13}N is a pair of this character. The normal states of ^{13}C and ^{13}N belong to the same T-multiplet, with $T=\frac{1}{2}$, the ^{13}C being the $T_\zeta=\frac{1}{2}$ and the ^{13}N the $T_\zeta=-\frac{1}{2}$ state of the T-multiplet.

Table 1

$M(-\frac{1}{2})-M(\frac{1}{2})$			$M(-\frac{1}{2})-M(\frac{1}{2})$			$M(-\frac{1}{2})-M(\frac{1}{2})$		
n	obs.	calc. (4)	n	obs.	calc. (4)	n	obs.	calc. (4)
1	−0·79	(−0·79)	15	2·94	2·82	29	5·02	4·99
3	−0·02	0·09	17	3·16	3·16	31	5·35	5·27
5	−0·07 ?	0·70	19	3·46	3·49	33	5·56	5·54
7	0·94	1·20	21	3·84	3·82	35	5·78	5·81
9	1·17	1·65	23	4·12	4·13	41	6·40	6·57
11	2·12	2·07	25	4·30	4·42			
13	2·40	2·45	27	4·90	4·71			

The assumptions made are supported by the fact that one obtains good agreement (cf. table 1) using formula (4), which contains only one arbitrary constant. A similar confirmation of (4) can be obtained by comparing the masses (Delsasso, White, Barkas and Creutz, 1940) of the normal states of ^{10}C and ^{10}Be (calc. 3·72, meas. 3·92). There are a number of cases in which one can arrive at an inequality concerning the masses. Thus, e.g., the normal state of ^{10}Be belongs, no doubt, to a T-multiplet with $T=1$. There must, therefore, also be a state of ^{10}B the mass of which differs from that of ^{10}Be by $-0{\cdot}79+\frac{1}{2}1{\cdot}27\times 8/10^{1/3}=1{\cdot}57$ mMu. This may be either the normal state of ^{10}B or an excited state, in which case the normal state lies *lower* (and belongs to a $T=0$ T-multiplet). The latter is the case, the mass difference between the normal states being −0·10 mMu. It is clear that a similar comparison can be made for all pairs of isobars (Wigner, 1937 b; Barkas, 1939).

Two assumptions are necessary for the derivation of (4): that concerning the electrostatic energy and that underlying the 2nd approximation, namely, the equality of proton-neutron, proton-proton and neutron-neutron interactions. Both assumptions seem to be well confirmed for the nuclei of table 1 and the ^{10}C–^{10}Be pair. Since there is nothing particular about these nuclei from the point of view of electrostatic energy—except that they all have relatively low

* Most experimental values are taken from a recent review on β decay by E. J. Konopinski. His value for $n=7$ (1·48 mMu.) differs considerably from that given by Haxby, Shoupp, Stephens and Wells (1940), and the latter was adopted in this case. For $n=1$ cf. any compilation of nuclear masses; for $n=3$ cf. O'Neal and Goldhaber (1940) and Brown (1941); for $n=5$ see the interesting discussion by Bloch (1940).

masses—we shall assume henceforth that our assumptions concerning the electrostatic energy are correct for light nuclei and low excitation. From the point of view of the 2nd approximation, however, it is well to remember that all the cases in which (4) could be quantitatively tested refer to pairs of isobars in which the number of protons in one isobar is equal to the number of neutrons in the other and vice versa. The relation (4) would follow for these from the assumption that the difference between proton-proton interaction and neutron-neutron interaction is given by the electrostatic energy (Feenberg and Knipp, 1935; Feenberg and Share, 1936; Fowler, Delsasso and Lauritsen, 1936; Feenberg and Wigner, 1937). Thus (4) would hold for these pairs even if the proton-neutron interaction were entirely different from the neutron-neutron or proton-proton interaction. The only data which permit a comparison of the former interaction with one of the latter are derived from the scattering experiments which form the basis of figure 1.

The natural limits of the 2nd approximation are: (*a*) very high atomic numbers, for which the next approximation to the electrostatic energy becomes appreciable, and (*b*) higher excitations at which the density of levels is so large that even a small interaction leads to an appreciable mixing of states. We shall see later that the limit (*a*) certainly lies at very high charges, if it is ever reached at all. It is unknown at present at what level densities the second cause for the breakdown of the 2nd approximation becomes operative.

§ 5. THE 1ST APPROXIMATION

In the 1st approximation the Hamiltonian can be written in terms of the space co-ordinates alone, entirely without the s or τ operators. There are two kinds of interaction which satisfy these requirements (and do not contain the second or higher powers of the velocity). These are the ordinary potential and the so-called Majorana exchange interaction (Majorana, 1933). The Majorana interaction between particles 1 and 2 will be denoted by M_{12}; it involves an interchange of the space co-ordinates

$$M_{12}\psi(r_1, \sigma_1, \tau_1, r_2, \sigma_2, \tau_2, r_3, \sigma_3, \tau_3, \ldots) = V_M(r_{12})\psi(r_2, \sigma_1, \tau_1, r_1, \sigma_2, \tau_2, r_3, \sigma_3, \tau_3, \ldots). \quad \ldots\ldots(6)$$

Here $V_M(r^{12})$ is an unspecified function of the distance of the first two particles. It is usually assumed that V_M has the shape of a reversed bell. Because of the antisymmetry of ψ, one can also write

$$M_{12} = -\tfrac{1}{4}[1+(\mathbf{s}_1 \cdot \mathbf{s}_2)]\,[1+(\boldsymbol{\tau}_1 \cdot \boldsymbol{\tau}_2)]\,V_M(r_{12}), \quad \ldots\ldots(6a)$$

where Dirac's well known relation is used for the interchange of spin co-ordinates (Dirac, 1935, § 61).

Just as the 2nd approximation predicts the coincidence of more ($(2T+1)(2J+1)$) energy levels than a rigorous calculation ($2J+1$), so the 1st approximation predicts even higher degeneracies than the 2nd approximation. This higher degeneracy will be resolved, of course, if one goes back to the second

approximation by introducing the spin-dependent forces. The situation is, in principle, quite similar to that encountered in ordinary spectroscopy where the $(2S+1)(2L+1)$-fold degenerate energy levels are resolved by the introduction of spin-dependent forces into $(2J+1)$-fold degenerate levels.

In detail, however, the situation is much more involved. Even if the proton-neutron, proton-proton, etc., forces were quite different one would have instead of the single S of spectroscopy two quantum numbers: the spin-angular moments S_P and S_N for proton and neutron are separately constants of motion (cf. chapter xvi of Wigner, 1931). The equality of proton-neutron, etc., interactions causes further coincidences of energy levels. A rigorous general treatment has been given by the methods of group theory (cf. Wigner, 1937 a; Hund, 1937 a), but will not be reproduced here.

§ 6. THE SUPERMULTIPLETS: TWO PARTICLES

We shall consider, first, as an example, a two-particle system.* This is mathematically the simplest case, but physically not a very favourable one, because the deviations from the 1st approximation become particularly significant in this case.

In the 1st approximation the Hamiltonian can be written and, therefore, also solved, using the space co-ordinates alone as variables. Some solutions will be symmetric, the others antisymmetric functions of the co-ordinates of the two particles. There will be a $(2L+1)$-fold degeneracy, corresponding to an azimuthal quantum number L. The symmetric wave functions would give, in ordinary spectroscopy, singlets; the antisymmetric ones triplets. In the present case, the former can be completed by multiplication with the following six functions of σ and τ to antisymmetric functions in the sense of (1 *a*):

$$\begin{aligned}&\delta(\tau_1,-1)\,\delta(\tau_2,-1)\,a(\sigma_1,\sigma_2), \quad s(\tau_1,\tau_2)\,a(\sigma_1,\sigma_2), \quad \delta(\tau_1,1)\,\delta(\tau_2,1)\,a(\sigma_1,\sigma_2),\\ &\qquad a(\tau_1,\tau_2)\,\delta(\sigma_1,1)\,\delta(\sigma_2,1),\\ &\qquad a(\tau_1,\tau_2)\,s(\sigma_1,\sigma_2), \qquad\qquad\qquad\qquad\qquad\qquad \ldots\ldots(7)\\ &\qquad a(\tau_1,\tau_2)\,\delta(\sigma_1,-1)\delta\,(\sigma_2,-1).\end{aligned}$$

Herein $a(\sigma_1,\sigma_2)$ and $a(\tau_1,\tau_2)$ are the antisymmetric functions of their arguments,

$$a(\tau_1,\tau_2)=2^{-\frac{1}{2}}[\delta(\tau_1,-1)\,\delta(\tau_2,1)-\delta(\tau_1,1)\,\delta(\tau_2,-1)], \qquad \ldots\ldots(7a)$$

while

$$s(\tau_1,\tau_2)=2^{-\frac{1}{2}}[\delta(\tau_1,-1)\,\delta(\tau_2,1)+\delta(\tau_1,1)\,\delta(\tau_2,-1)]. \qquad \ldots\ldots(7b)$$

Hence, every symmetric function of the space co-ordinates gives rise to *six* states of the nuclei containing two particles. Two of these, corresponding to the first and last column of (7), are states of a system containing two protons and two neutrons, respectively (cf. (2) and (2*a*)). Multiplication of the symmetric function of the space co-ordinates with one of the functions of the central column

* The development of § 6 and § 7 follows that given in Wigner, 1937 a.

of (7) gives a state of the deuteron. The dependence on σ_1, σ_2 shows that the first of these is a singlet, the others form a triplet.

An antisymmetric function of the space co-ordinates can be completed to an antisymmetric function of all co-ordinates by multiplication with ten different functions of τ_1, τ_2, σ_1, σ_2, but it is needless to go into the details of this case.

All the six states, derived from one symmetric function of the space co-ordinates by multiplication with one of the six functions (7), have the same energy in the 1st approximation. This is a very crude approximation in the case of two particles, since it is known that the energy of the triplet state of ^{2}H differs by as much as $2\frac{1}{2}$ mMu. from the singlet state. However, the 1st approximation is a good deal more accurate for heavier nuclei—although the above discussion would require much more space in these cases.

If one introduces the spin-dependent forces as a perturbation, the six-fold degenerate state will split up. However, if the perturbation is " symmetric ", the 2nd approximation will remain valid and the binding energy of the three

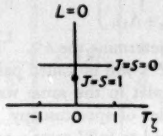


Figure 3*a*.

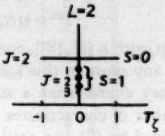


Figure 3*b*.

Figure 3. The (100) supermultiplet for $L=0$ (figure 3*a*) and for $L=2$ (figure 3*b*).

states corresponding to the first line of (7) will remain the same. The spin-dependence of these three states corresponds to a singlet ($S=0$), and their J will be, therefore, equal to the L of the space co-ordinate wave function from which they originate. The dependence of the remaining three functions of (7) on σ corresponds to a triplet state. Hence, apart from the case $L=0$, these will give rise to three energy levels with the J values $L-1$, L and $L+1$. For $L=0$—this is actually the only important case for the deuteron—the last three functions of (7) give only one energy level, the J of which is 1. If the spin-dependent forces are very weak, the wave function of this state will be simply the space co-ordinate wave function, multiplied with one of the functions (7). In this case it will have all the characteristics of a $^3S_1$ state. However, if the spin-dependent forces are strong, the spin-angular momentum and the orbital-angular momentum may not be good quantum numbers.*

The level scheme of the above supermultiplet in 2nd approximation is shown in figure 3*a* for an S-term, in figure 3*b* for a term with $L\neq 0$.

* The validity of the quantum numbers L and S in the two-particle problem has been discussed by Bethe (1940) and by Eisenbud and Wigner (1941).

§ 7. THE SUPERMULTIPLETS: ARBITRARY NUMBER OF PARTICLES

The situation is very similar in the case of more than two particles, with the difference, however, that the symmetry properties of the wave function containing only the space co-ordinates can be much more involved. They can be characterized (cf. Wigner, 1937a; Hund, 1937a) by four integer "partition quantum numbers" $\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4$, for which

$$\Lambda_1 \leqslant \Lambda_2 \leqslant \Lambda_3 \leqslant \Lambda_4 ; \qquad \Lambda_1+\Lambda_2+\Lambda_3+\Lambda_4 = n. \qquad \ldots\ldots(8)$$

In ordinary spectroscopy there are only two Λ's, viz. Λ_1 and Λ_2, and $\Lambda_1 \leqslant \Lambda_2$; $\Lambda_1+\Lambda_2=n$ take the place of (8). This is connected with the fact that in ordinary spectroscopy there is only one spin, with only two values, while we have two spins, σ and τ, with four values altogether. In ordinary spectroscopy, the character of the multiplet is determined if $\frac{1}{2}(\Lambda_2-\Lambda_1)=S$ is given. In our case, in order to characterize the multiplet, we need three numbers,

$$\left.\begin{aligned} P &= \tfrac{1}{2}(\Lambda_4+\Lambda_3-\Lambda_2-\Lambda_1), \\ P' &= \tfrac{1}{2}(\Lambda_4-\Lambda_3+\Lambda_2-\Lambda_1), \\ P'' &= \tfrac{1}{2}(\Lambda_4-\Lambda_3-\Lambda_2+\Lambda_1), \end{aligned}\right\} \qquad \ldots\ldots(9)$$

which, together with n (cf. (8)), completely determine the Λ's. The Λ's will not be mentioned any more in this article, the P's will be called partition quantum numbers. They characterize a supermultiplet in the same way as the single quantum number S characterizes a multiplet of spectroscopy. The numbers P, P', P'', derived by (9), can easily be proved to fulfil certain conditions which will play a major rôle in later developments. These are:—

(*a*) They are all integers if n is even, all half-integers if n is odd (cf. the similar rule for S).

(*b*) Their sum is even if n is a multiple of 4; it is odd if n is even but not a multiple of 4. In general, it has the form

$$P+P'+P'' = 2k-\tfrac{1}{2}n, \qquad \ldots\ldots(10)$$

where k is an integer (Λ_4).

(*c*) They satisfy the inequalities

$$P'+P'' \geqslant 0; \quad P \geqslant P'; \quad P' \geqslant P'' \qquad \ldots\ldots(10a)$$

It follows from (10*a*) that only P'' can be negative, and that its absolute value is smaller than that of P'. In addition, for a given number n of particles, only those multiplets are possible for which $P+P' \leqslant P''+n$.

The formulae which give the terms into which a ($P\,P'\,P''$) supermultiplet splits under the influence of spin-dependent forces are very complicated and will not be given here. Instead, table 2 gives for the most important supermultiplets the S–T combination which they contain. For $L=0$, i.e., for S terms, these are the J–T combinations which that supermultiplet gives if one goes over to the 2nd approximation. For $L \neq 0$, one obtains the J values, associated with a given T, by taking the possible vector sums of L and the S

which appears associated with the T in question in table 2. Figure 3 illustrates this for the supermultiplet (1 0 0).

Table 2

$(P\ P'\ P'')$	S	T	$(P\ P'\ P'')$	S	T	$(P\ P'\ P'')$	S	T	$(P\ P'\ P'')$	S	T
(0 0 0)	0	0	(1 0 0)	0	1	(3 0 0)	0	1	$(\frac{3}{2}\ \frac{3}{2}\ \pm\frac{1}{2})$	$\frac{1}{2}$	$\frac{1}{2}$
(1 1 0)	0	1		1	0		1	0		$\frac{1}{2}$	$\frac{3}{2}$
	1	0	(1 1 $\pm$1)	0	0		1	2		$\frac{3}{2}$	$\frac{1}{2}$
	1	1		1	1		2	1		$\frac{3}{2}$	$\frac{3}{2}$
(2 0 0)	0	0	(2 1 0)	0	1		0	3	$(\frac{5}{2}\ \frac{1}{2}\ \pm\frac{1}{2})$	$\frac{1}{2}$	$\frac{1}{2}$
	1	1		1	0		3	0		$\frac{1}{2}$	$\frac{3}{2}$
	2	0		1	1	$(\frac{1}{2}\ \frac{1}{2}\ \pm\frac{1}{2})$	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{3}{2}$	$\frac{1}{2}$
	0	2		1	1	$(\frac{3}{2}\ \frac{1}{2}\ \pm\frac{1}{2})$	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{3}{2}$	$\frac{3}{2}$
(2 1 $\pm$1)	1	0		0	2		$\frac{1}{2}$	$\frac{3}{2}$	$(\frac{5}{2}\ \frac{1}{2}\ \pm\frac{1}{2})$	$\frac{1}{2}$	$\frac{1}{2}$
	0	1		2	0		$\frac{3}{2}$	$\frac{1}{2}$		$\frac{5}{2}$	$\frac{1}{2}$
	1	1		1	2	$(\frac{3}{2}\ \frac{3}{2}\ \pm\frac{1}{2})$	$\frac{1}{2}$	$\frac{1}{2}$			
	1	2		2	1		$\frac{3}{2}$	$\frac{3}{2}$			
	2	1									

There are two rules which are worth remembering:—

(i) If an S–T combination is contained in a supermultiplet, the T–S combination is contained equally often. This is natural, since S and T play, from a mathematical point of view, the same rôles in the 1st approximation, just as σ and τ do.

(ii) The highest T is equal to P, and the highest S with which it occurs is equal to P'; it follows then from (i) that the highest S is also P, and the highest T with which it occurs is P'. It follows from this, e.g., that only those supermultiplets give terms for a nucleus with a given T_ζ, the P of which is at least as great as T_ζ.

If the 1st approximation is good, the azimuthal quantum number L, the total angular momentum J and the parity should be good quantum numbers in addition to the P, P', P'', which are also good quantum numbers. If, furthermore, —as we shall always assume—the most important perturbation is still symmetric, i.e. if the 2nd approximation is sufficiently good to calculate the correct linear combinations of the states into which the supermultiplet splits, T also will be a

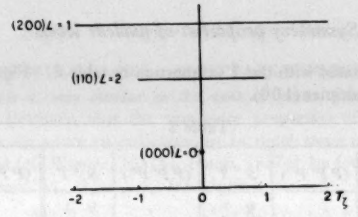


Figure 4*a*.

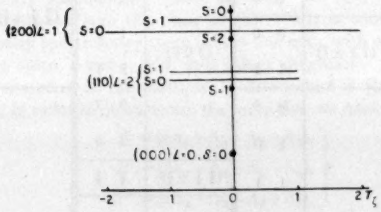


Figure 4*b*.

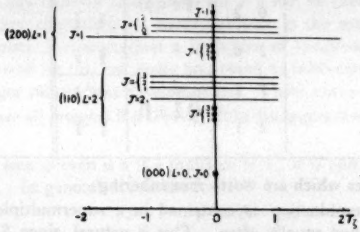


Figure 4*c*.

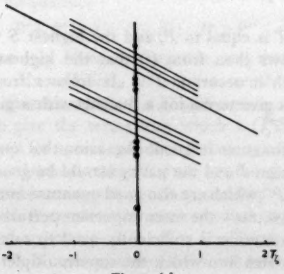


Figure 4*d*.

good quantum number. S will be a good quantum number only if there is only one S associated in the supermultiplet with the T in question. This is true in the most important cases.

However, if the spin-dependent forces are large, they will mix all the quantum numbers of terms which are close to each other. P, P', P'', S and, in particular, L will cease to be good quantum numbers. This seems to be the situation apart from a few exceptional cases. We shall not discuss the experimental material at this point, since much of the rest of this *Report* will be devoted to its investigation. Figure 4 shows the transition from 1st to 2nd and last approximation in the case of a few typical multiplets. Figure 4*a* shows a (0 0 0) supermultiplet with $L=0$, a (1 1 0) supermultiplet with $L=2$ and a (2 0 0) supermultiplet with $L=1$. So many states coincide in some cases, even for one nucleus (one T_z), that the degeneracy has not been indicated in this figure. Figures 4*b* and 4*c* correspond to the 2nd approximation. In the former, the splitting of the supermultiplet into terms with different T is taken into account; not, however, the interaction between spin and orbit. This causes further splittings, which are illustrated in figure 4*c*. Figure 4*b* corresponds to a highly artificial situation: actually the same interactions which cause the splittings illustrated in figure 4*b* also cause those of figure 4*c*. The degeneracy in figure 4*b* is $(2S+1)(2L+1)$, in figure 4*c* it is $2J+1$. The relative position of the terms into which a supermultiplet splits is purely arbitrary in figures 4*b* and 4*c*. Figure 4*d* differs from 4*c* only by the slight tilt which the Coulomb forces give to the T-multiplets of figure 4*c*.

§ 8. QUALITATIVE DISCUSSION OF THE POSITION OF LOW SUPERMULTIPLETS

Before giving detailed calculations on the position of supermultiplets in the energy scale, it may be useful to give a qualitative outline of the results to be obtained.

It is evident from figures 2, especially 2*c*, that the normal state of a stable nucleus is always the highest member of its T-multiplet, with $T_z=T$. One sees, furthermore, from figures 4, especially 4*d*, that unless the slope of the line connecting the members of a T-multiplet is very small, or the splitting of the supermultiplet very large, the stable isobar will have the highest T possible in the supermultiplet.

Figure 4. Transition from 1st to 2nd and last approximation. Figure 4*a* represents three supermultiplets with azimuthal quantum numbers L given at the left, in 1st approximation. Figures 4*b* and 4*c* represent the splitting of these supermultiplets in 2nd approximation Figure 4*d* takes the electrostatic interaction into account.

Now, for mass numbers which are multiples of four, the stable isobar has the same number of neutrons and protons up to $n=36$. This shows that the highest T is zero in the lowest-lying supermultiplet, i.e. the lowest supermultiplet is (0 0 0). For odd mass numbers, the T_ζ of the stable isobar is $\frac{1}{2}$, up to $n=35$. This shows again that the highest T of the lowest supermultiplet is $\frac{1}{2}$ and the supermultiplet itself is $(\frac{1}{2}\ \frac{1}{2}\ \pm\frac{1}{2})$ (cf. table 2 or rules (ii) and (*c*), § 7). It is seen that in these two cases, in the lowest supermultiplet, P, P' and P'' are as small as compatible with the mass number n. We shall later generalize this rule to give the approximate position of the lowest term of a supermultiplet *in the 1st approximation*,

$$E(P, P', P'') = A(n) + \tfrac{1}{2}L(n)[P^2 + 4P + P'^2 + 2P' + P''^2], \quad \ldots\ldots(11)$$

where $A(n)$ and $L(n)$ are smooth functions of the mass number n.

The situation is particularly interesting for even nuclei the mass numbers of which are *not* multiples of 4. Rule (*b*) of § 7 shows that the lowest supermultiplet cannot be (0 0 0); it is (1 0 0). Table 2 shows that this gives two T-multiplets; a triplet with $T=0$ and a singlet with $T=1$. It is, therefore, *a priori* impossible to tell which would be the stable element in the 2nd approximation, the $T_\zeta=0$

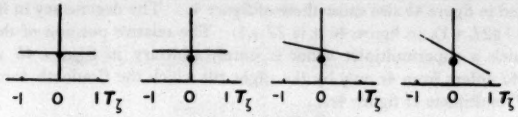


Figure 5*a*. Figure 5*b*. Figure 5*c*. Figure 5*d*.

Figure 5. The (100) supermultiplet for $L=0$ in 1st approximation (figure 5*a*), in 2nd approximation (figure 5*b*) and deformed by the electrostatic interaction for low n (figure 5*c*) and high n (figure 5*d*).

or the $T_\zeta=1$ (cf. figures 5*b* and 5*c*). Experience tells us that for very low n, up to $n=14$, the $T_\zeta=0$ is the stable element, so that we must assume that the spin-dependent forces shift the $S=1$, $T=0$ term down, the $S=0$, $T=1$ term up.

We see, however, that from $n=18$ (to $n=46$) $T_\zeta=1$ holds for the stable isobar. It would be possible, therefore, to assume that in these cases the $S=0$, $T=1$ term is lower in the 2nd approximation than the $S=1$, $T=0$ term. It is more reasonable, however, to assume that the relative position of the T-multiplets does not change from $n=14$ to $n=18$ but that the (negative) slope of the line connecting the members of the T-multiplet has increased so much that although in 2nd approximation the $T=0$ is the lowest T-multiplet, the electrostatic interaction causes the $T_\zeta=1$ term of the $T=1$ T-multiplet to lie lower than the $T=0$ term of the $T=0$ T-multiplet. Indeed, (4) shows that the (negative) slope of the line connecting the members of a T-multiplet increases with increasing n—which is only another expression for the fact that the electrostatic potential of the ^{18}F nucleus is higher than that of the ^{14}C nucleus. The situation is illustrated in figure 5. Figure 5*a* corresponds to the 1st, figure 5*b* to the 2nd

approximation and figures 5*c* and 5*d* take the electrostatic forces also into account. The former applies for $n \leqslant 14$, the latter for $n \geqslant 18$. A definite proof that, for $n=18$, the $T=0$ T-multiplet is lower in the 2nd approximation than the $T=1$ T-multiplet, can be obtained by calculating, by means of (4), the $T_\zeta=0$ member of the T-multiplet, of which the $T_\zeta=1$ member is the normal state of ^{18}O. Such a calculation has been carried out, for $n=10$, below table 1 in § 4. Table 3 gives the mass differences (in mMu.) of the lowest $T=0$ and $T=1$ T-multiplets calculated in this way. These differences are understood as the splitting of the lowest supermultiplet (1 0 0) due to spin-dependent forces in the 2nd approximation.

Table 3

n	6	10	14	18	22	26	30
Spin energy	4·6	2·20	2·45	1·25	2·0	1·25	1·05

It is seen that, on the whole, the effect of spin-dependent forces decreases with increasing number of particles but has the same sign throughout ($S=1$, $T=0$ lower than $S=0$, $T=1$).

If the above interpretation of the normal state of light nuclei is correct, all even-even nuclei (nuclei which contain an even number of protons and of neutrons) are in singlet states, at least as long as the 1st approximation is adequate for calculating the wave functions. The rule that the J of all even-even nuclei is zero then shows that the orbital angular momentum L is zero for even n in the lowest supermultiplet. It then follows further that $J=1$ should hold for the low $T=0$, $S=1$ T-multiplet for $n=2$, 6, 10, 14. Experimentally, the J values of the normal states of ^{2}H, ^{6}Li, ^{10}B, ^{14}N are known; they are 1, and thus seem to support the above theory. This point will be further discussed in § 10.

Thus far, we have succeeded only in determining the character of the lowest supermultiplets of the light elements. Beyond $n=40$, the number of neutrons exceeds the number of protons by more than 1 in the stable isobar even if the mass number is odd. Again this could be interpreted by assuming that beyond 40, the $(\frac{3}{2}\ \frac{1}{2} \pm \frac{1}{2})$ multiplet is lower than the $(\frac{1}{2}\ \frac{1}{2} \pm \frac{1}{2})$ multiplet even in the 1st approximation. It appears more reasonable, however, to assume that the $(\frac{1}{2}\ \frac{1}{2} \pm \frac{1}{2})$ multiplet is always lowest in odd elements in the 1st approximation, but the effect of the electrostatic interaction becomes so big in the region around $n=40$, that they push a $T_\zeta=\frac{3}{2}$ member of the higher $(\frac{3}{2}\ \frac{1}{2} \pm \frac{1}{2})$ multiplet below all the members of the $(\frac{1}{2}\ \frac{1}{2} \pm \frac{1}{2})$ multiplet. This view can be supported by calculating the position of the $T_\zeta=\frac{1}{2}$ member of the $T=\frac{3}{2}$ T-multiplet and comparing it with the position of the normal state for $T_\zeta=\frac{1}{2}$. We have, e.g., for $n=43$ the stable $^{43}_{20}$Ca and the radioactive $^{43}_{21}$Sc, which emits a positron with a maximum energy of 1·40 mMu. In order to obtain the mass difference between ^{43}Sc and ^{43}Ca, the mass 0·548 of the electron has to be added twice to the 1·40 mMu. (in case of electron emission no such correction is necessary). Thus the mass ofthe ^{43}Sc atom is greater than that of the ^{43}Ca atom by 2·50 mMu.

We can calculate further from (4) that the $T_\zeta = \frac{1}{2}$ member of the T-multiplet lies higher than the $T_\zeta = \frac{3}{2}$ member by 6·45 mMu. Thus the state of ^{43}Sc which is in the T-multiplet of the normal state of ^{43}Ca has a mass which is by 6·45 mMu. higher than the ^{43}Ca mass, i.e., $6{\cdot}45 - 2{\cdot}50 = 3{\cdot}95$ mMu. higher than the mass normal state of ^{43}Sc. A comparison with the numbers in table 3 shows that of the this is a good deal more than the splitting within a supermultiplet would be for an n as high as 43.

One can calculate in the above fashion the mass difference of the states with $T_\zeta = T = \frac{1}{2}$ and the state $T_\zeta = \frac{1}{2}$, $T = \frac{3}{2}$. Table 4 gives the results for those n for which the experimental data are available.

Table 4

n	23	27	29	31	35	41	43	47
ΔM	8·10	7·40	7·25	6·80	5·55	5·00	3·95	3·80

One sees that the ΔM lie close to a smooth curve $(190/n)$ and there is no discontinuity between the range ($n<37$), where the $T_\zeta = \frac{1}{2}$, and the range ($n>39$), where the $T_\zeta = \frac{3}{2}$ elements are stable. Since in the former case the normal states of the isobars certainly belong to different supermultiplets, this must hold also in the latter case.

Table 4 gives directly the distance between the lowest T-multiplet of the $(\frac{3}{2}\ \frac{1}{2} \pm \frac{1}{2})$ and of the $(\frac{1}{2}\ \frac{1}{2} \pm \frac{1}{2})$ supermultiplets in the 2nd approximation: it is about $190/n$. Since the distance between the T-multiplets of the same supermultiplet, as given by table 3, is much smaller (about $30/n$), this is also the order of magnitude of the distance between different supermultiplets in the 1st approximation.

We shall adopt the above picture throughout the whole periodic table. We shall assume that, in the 1st approximation, the energy of the lowest $(P\ P'\ P'')$ supermultiplet runs parallel to the magnitude of the numbers P, P', P''. This qualitative statement is formulated more quantitatively in (11), the $L(n)$ of which is therefore positive, although it decreases with increasing n. As a consequence, in the 1st approximation the lowest state of an isobar with higher $|T_\zeta|$ will be higher than the lowest state of an isobar with a lower $|T_\zeta|$ (this is the order of the supermultiplets in figure 4*a*). The spin-dependent forces, it will be assumed, do not alter the energy levels substantially (in fact, they play a major rôle only in the elements discussed in connection with table 3). However, with increasing n, the electrostatic forces increase greatly in importance. The lines connecting the members of a T-multiplet become, according to (4), steeper and steeper. Finally, they become so steep that the right side of a higher supermultiplet will reach lower than the shorter right side of a lower supermultiplet and the T_ζ of the stable isobar will become higher than for the preceding n. This process, schematically represented in figure 6, will continue and, by the end of the periodic table, the T_ζ of the stable isobar will assume values as high as 20. This explanation of the shift to higher supermultiplets

with increasing n is very similar to the shift te tho higher T-multiplet, illustrated in figure 5.

It may be superfluous to remark that the above explanation of the increase in T_ζ, i.e. of the preponderance of neutrons in heavy elements, does not go, in its content, much beyond Heisenberg's ideas. It would, perhaps, be more useful

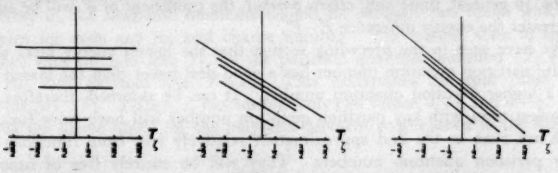


Figure 6*a*. Figure 6*b*. Figure 6*c*.

Figure 6. Increase of the isotopic spin of the stable isobar with increasing n, i.e. increasing (negative) slope of the lines connecting the members of a T-multiplet.

to answer the question which the reader may put at this point: how accurate can the 1st approximation be if the corrections thereto (particularly those due to the electrostatic interaction) so fundamentally upset the order of the terms? This question will be discussed in the next sections.

§ 9. DIRECT EVIDENCE CONCERNING THE APPLICABILITY OF THE 1ST APPROXIMATION

There are three direct sources of information on the accuracy of the 1st approximation: (*a*) magnetic moments, (*b*) quadrupole moments, (*c*) lifetimes of β-active substances.

We can imagine that the total Hamiltonian is divided into three parts: H_0, H_1, H_2. The H_0 contains only that part of the interaction which is independent of both ordinary and isotopic spin, H_1 is that part of the interaction which depends on ordinary spin, but not on isotopic spin, while H_2 is the electrostatic interaction which depends also on the isotopic-spin coordinates. Considering H_0 alone corresponds to the 1st approximation, the 2nd approximation corresponds to the Hamiltonian H_0+H_1, and the correct wave functions are the characteristic functions of $H_0+H_1+H_2$.

For the characteristic functions of H_0 all the quantum numbers P, P', P'', L, T, S, J are good quantum numbers. If we introduce H_1 as a perturbation, the above will be good quantum numbers only if we assume that the perturbation is weak. Otherwise, the characteristic functions of H_0 will suffer perturbations and become approximately

$$\psi \to \phi = \psi + \Sigma'(\psi', H_1\psi)(E-E')^{-1}\psi', \qquad \ldots\ldots(12)$$

where ψ is one characteristic function of H_0 with the energy parameter E, and the summation runs over all other characteristic functions ψ' of H_0. The energy parameter of ψ' is denoted by E'.

Since T and J remain good quantum numbers even if H_1 is taken into account, the coefficient of ψ' in (12) must vanish if the T and J of ψ' are different from the T and J of ψ. This is indeed the case: the matrix element of H_1 connecting ψ' and ψ is zero unless the total isotopic spin T and the total angular momentum J have the same values for ψ' and ψ. If this condition is satisfied, $(\psi', H_1\psi)$ will be, in general, finite and, *ceteris paribus*, the coefficient of ψ' will be smaller the greater the energy difference $E-E'$.

We have seen in the preceding section that the lowest energy level with a definite partition quantum number lies a good deal lower than the lowest term with a higher partition quantum number. It can be assumed, therefore, that the lowest term with any partition quantum number will have wave functions which are even in the 2nd approximation relatively free from functions with *higher* partition quantum numbers. They will be entirely free of functions with lower partition quantum numbers if no lower partition quantum number gives a T as high as that of the term in question. It follows from this that P, P', P'' will be good quantum numbers even in the 2nd approximation for the normal state of not-too-heavy nuclei. Let us consider, for example, the normal state of a nucleus with isotopic spin T_ζ, the partition quantum numbers of which are $(T_\zeta\,0\,0)$. It will belong to a T-multiplet with $T=T_\zeta$. Its wave function in the 2nd approximation, ϕ, will contain no ψ', the P of which is less than T_ζ, because such a ψ' would belong (cf. rule (ii), §7) to a T-multiplet with $T\leqslant P$, and hence $T<T_\zeta$ and would be the state of another nucleus. Similarly, ϕ will contain the ψ' for which $P>T$, only with very small coefficients because the energy E' of such a ψ' is much higher than the energy E of ψ (cf. (11)). For the same reason even the coefficients of the ψ' with partition quantum numbers $(T_\zeta, 1, 1)$, $(T_\zeta, 2, 0)$, etc., will be small. But, because the energy differences are relatively small, these coefficients will be larger than those associated with the ψ', the partition quantum numbers of which differ greatly from those of ψ.

The situation will be very different for the more highly excited states. If they are closely interspersed with states of other partition quantum numbers in the 1st approximation, H_1 may cause such an intermingling among them that it is quite impossible to assign partition quantum numbers in the 2nd approximation. The same will be true in many cases even for the lowest term for the azimuthal quantum number L, because the states with other L will not lie, in general, much higher than the normal state. The same would be true also for S, except that in most cases the partition quantum numbers and T together uniquely determine S. This may go even further: the S, together with J, which is always a good quantum number, may determine L uniquely (if $S=0$) or within a double alternative (if $S=\frac{1}{2}$).

The effect of the electrostatic interaction H_2 on the validity of the 2nd approximation will be discussed later. It will be found that, in spite of its enormous magnitude and effect on the energy values, it does not influence the wave functions greatly.

§ 10. MAGNETIC MOMENTS

The magnetic moments of odd elements were discussed from the above point of view by Margenau and Wigner (1939). However, in order to calculate the total magnetic moment they had to adopt a plausible, but by no means rigorously justifiable, assumption first employed by Way (1939). They conclude that, for every J, the magnetic moments should lie between two limits which are different for even and for odd charge number. The normal state of an odd element with isotopic spin T_ζ has the partition quantum numbers $(T_\zeta, \frac{1}{2}, \pm\frac{1}{2})$ and the total isotopic spin $T = T_\zeta$. This determines the spin value $S = \frac{1}{2}$ (cf. rule (ii), § 7). Hence, L can assume only two values, $J + \frac{1}{2}$ and $J - \frac{1}{2}$. The first of the above limits for the magnetic moment corresponds to the first, the

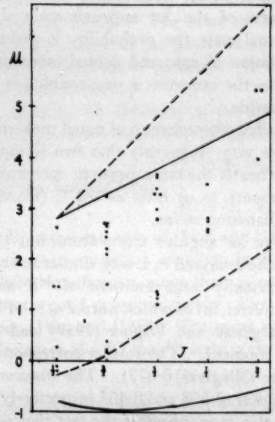


Figure 7*a*.

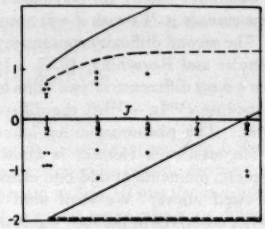


Figure 7*b*.

Figure 7. Magnetic moments μ of nuclei with odd number of protons (figure 7*a*) and odd number of neutrons (figure 7*b*) as functions of the total mechanical moment J.

second to the second alternative. Actually, however, the wave function can be a linear combination of a ψ with $L = J + \frac{1}{2}$ and a ψ' with $L = J - \frac{1}{2}$, and the magnetic moments must be expected to lie between the two limits, represented by full lines in figures 7*a* and 7*b*.

This expectation is fulfilled in all (about 50) cases except four—^{113}In, ^{115}In, ^{83}Kr ?, ^{87}Sr.

This seems to give a very favourable picture of our assumptions. There are, however, two facts which considerably blur that picture. The first one is that an even simpler picture than our own gives limits for the magnetic moments which

are as close to each other as are our limits and which also include all moments, without exception. These limits are Schmidt's (Schmidt, 1937; Schuler, 1937) (broken lines on figures 7*a* and 7*b*), and can be interpreted by assuming that the wave function of the normal state is a linear combination of two wave functions. The first of these is an even state of the preceding stable nucleus with $J=0$, coupled with a state of the extra particle with orbital momentum $l=J-\frac{1}{2}$. In the second state the preceding nucleus has odd parity and the orbital momentum of the extra particle is $J+\frac{1}{2}$. This picture is not at all unreasonable and gives limits for the momenta which are, as mentioned above, always obeyed. The fact that both Schmidt's limits and those of the 1st approximation are observed by most elements has as a consequence that, with four exceptions, all moments lie between narrower limits than either of the two pictures demands. From the point of view of the 1st approximation this requires the interpretation that in the normal state the probability is greater for a parallel than for an antiparallel orientation of spin and orbital momenta. In Schmidt's theory the conclusion would be the opposite, a result which is in some cases (e.g. ^{7}Li with $J=\frac{3}{2}$) not very plausible.

The second difficulty consists in the so-called phenomenon of equal moments (Schuler and Korsching, 1937). It happens very frequently that two isotopes with a mass difference of two units have very nearly the same magnetic moments; in one case (^{113}In$-^{115}$In) the difference amounts to as little as 0·2 % (Hardy, 1941). This phenomenon has found no explanation so far.

No even-even element is known to have an angular momentum, but the magnetic momenta of odd-odd elements can be analysed in a way similar to that discussed above. We shall omit the abnormally large moment of ^{40}K and restrict ourselves to the four stable odd-odd nuclei, all of which have $J=1$. The same assumptions which were used by Margenau and Wigner (1939) lead to the limits 0·855 and 0·323 for the magnetic moment. The former corresponds to a pure ^{3}S, the latter to a pure ^{3}D state (^{3}P gives 0·677). The observed moments for ^{2}H, ^{6}Li, ^{10}B and ^{14}N are 0·856, 0·821, 0·598 and 0·403 respectively.* This is in agreement with the theory, but it also re-emphasizes the fact that our picture does not yield all the information that is desired. In the case under discussion, it fails to explain the gradual decrease of the magnetic moment with increasing n which is, probably, due to the gradual increase of the coefficient of the ^{3}D wave function in the actual normal state (Millman, Kusch and Rabi, 1939). Similarly, in the case of odd elements, our picture gave only a rather wide interval within which the magnetic moment should lie. In order to arrive at more precise predictions one must make use—as in so many other cases—of a more detailed model than ours (Rose and Bethe, 1937; Inglis, 1938 a, 38 b, 39; Bethe, 1938 a; Sachs, 1939; Phillips, 1940).

* For a list of references see Millman and Kusch, 1941.

§ 11. QUADRUPOLE MOMENTS

The quadrupole moment is usually defined as the average value of the quantity $2z^2+x^2+y^2$ for all protons in that state in which the angular momentum is parallel to the z axis. Expressed by means of the formalism of the isotopic spin it is

$$(\psi_{M=J},\ \Sigma_i \tfrac{1}{2}(1-\tau_{3i})(2z_i^2-x_i^2-y_i^2)\psi_{M=J}). \qquad \ldots\ldots(13)$$

It is usually assumed that the spatial distribution of the neutrons is the same as that of the protons, and thus a positive quadrupole moment corresponds to a prolate, a negative quadrupole moment to an oblate ellipsoidal shape of the nucleus.

The most important among the quadrupole moments is that of the deuteron. It was first measured by Rabi and his collaborators (Kellogg, Rabi, Ramsey and Zacharias, 1940).* The mere existence of this quadrupole moment shows that the normal state of the deuteron is not a pure ^{3}S state since such a state would have no quadrupole moment. It must be a linear combination of a ^{3}S and a ^{3}D wave function.†

Prior to the discovery of the deuteron's quadrupole moment it was assumed that H_1 consists essentially of the so-called Heisenberg force

$$-\tfrac{1}{2}[1+(\tau_1\,.\,\tau_2)]V(r_{12})=P_{12}Q_{12}V(r_{12}) \qquad \ldots\ldots(14a)$$

and the Bartlett force

$$\tfrac{1}{2}[1+(\sigma_1\,.\,\sigma_2)]V(r_{12})=Q_{12}V(r_{12}). \qquad \ldots\ldots(14b)$$

If H_1 consisted only of (14a) and (14b), the orbital angular momentum L would be a good quantum member in the 2nd approximation (which contains the whole Hamiltonian, since there is only one proton in ^{2}H). Thus the fact that L is not a good quantum number shows that H_1 contains terms which are different from (14a) and (14b). The most simple form of interaction which causes the admixture of a ^{3}D to a ^{3}S state is the so-called tensor interaction,‡

$$[3(\mathbf{r}\,.\,\sigma_1)(\mathbf{r}\,.\,\sigma_2)/r^2-(\sigma_1\,.\,\sigma_2)]V_t(r). \qquad \ldots\ldots(15)$$

It is simplest to assume that $V_t(r)$ has the general shape of a reversed bell, with a minimum at $r=0$, from which it rises steadily to 0. The value of the minimum is called the depth, the distance r at which $V_t(r)$ rises to $\frac{1}{2}V_t(0)$ is called the half-width of the interaction. The following considerations (Rarita and Schwinger, 1941 a, 41 b) are based on the assumption that $V_t(r)$ has this simple shape.

If one assumes a reasonable form for the ^{3}D part of the wave function, the knowledge of the quadrupole moment permits one to calculate the probability (square of the coefficient) of the ^{3}D state. The estimates for this (Bethe,

* For comparison with measurements of the magnetic moment of the neutron, see Alvarez and Bloch (1940), and Rarita and Schwinger (1941a).

† All $^3P_1$ and $^1P_1$ states of a two-particle system are odd.

‡ This kind of interaction was suggested by Yukawa's meson theory of nuclear forces. Cf. Frohlich, Heitler and Kemmer (1938); Bhabha (1938); Kemmer (1938); Yukawa, Sakata and Taketani (1938); Bethe (1939, 40).

1939, 40; Rarita and Schwinger, 1941 a, 41 b) range between 4 % and 6 %. From this, again, it is possible to calculate the magnitude of the tensor force, i.e. the depth, $-V_t(0)$ of the $V_t(r)$. The result appears to be unique, although it depends somewhat on the assumed half-width, $2{\cdot}8 \times 10^{-13}$ cm., of $V(r)$. It is of the order of 11·5 mMu. This must be compared with the depth of the interaction which appears in the 1st approximation, and which is of the order of 15 mMu. (Breit, Thaxton and Eisenbud, 1939).

The partition quantum number is always a good quantum number in a two-particle system because it is given by the parity. For ^{3}H and ^{4}He, the above determined magnitude of the tensor interaction gives about 4 % for the amount of admixture of higher states to the wave function, determined by the 1st approximation. The admixture (Gerjuoy and Schwinger, 1941) consists in these cases of states with higher partition quantum numbers. The discussion of the magnetic moments shows that the total modification may be considerably greater than 4 % in heavier elements. However, it consists mostly of states with the same partition quantum number as ψ's. In view of the magnitude of the tensor interaction, it is quite remarkable that it causes so little admixture of states with higher partition quantum numbers to the normal state. The same holds true, apparently, not only for the tensor interaction but for other spin- or isotopic spin-dependent interactions as well. When reviewing their first calculations on the nuclei between He and O, which were based on an isotopic spin-dependent Hamiltonian, the authors noticed that the wave functions obtained differed only little from the wave functions obtained later by Feenberg and Phillips, who used a symmetric Hamiltonian (cf. § 15). The same holds true for He, for which one obtains an almost symmetric wave function even if one assumes that there is no proton-proton or neutron-neutron interaction (cf. Wigner, 1933). One can find numerous, even more striking, examples in the literature. The reason for this behaviour is, presumably, the magnitude of the energy difference between states with different partition quantum numbers in 1st approximation. It is probable, however, that the situation changes considerably at higher n.

It should be possible to give a general estimate for the quadrupole moments of heavier elements, using the above magnitude of the tensor force and the picture to be developed later (in § 14). However, no work has been done along this line so far.

§ 12. LIFETIMES OF NUCLEI EMITTING BETA RAYS

The calculation of the magnetic moment due to the orbital motion of the protons was possible only by making a rather crude assumption. No such assumption is necessary for the calculation of the lifetimes of β-active substances if the transition is an allowed one. On the other hand, there are at least three different formulations of Fermi's theory of β disintegration (Fermi, 1934) which predict different values for the transition probability between two states.

Fermi's theory, and also its results, show a great similarity to the theory of light emission. The transition probability for allowed transitions is a product of two factors. The first factor $I(w)$, depends on the energy of disintegration, wmc^2 (m is the electronic mass), and on Fermi's constant of interaction g,

$$I(w) = g^2 f(w). \qquad \ldots\ldots(16)$$

$$f(w) = (w^2-1)^{\frac{1}{2}}(w^4/30 - 3w^2/20 - 2/15) + \tfrac{1}{4}w \ln (w + (w^2-1)^{\frac{1}{2}}). \qquad \ldots\ldots(16a)$$

It is analogous to the expression $4e^2w/3mch^2$ occurring in the theory of optical transition probabilities. The second factor depends on the wave functions ψ_i and ψ_f of initial and final state. It is the sum of the squares of a few matrix elements (of the three components of the linear momentum in the optical case). The three above-mentioned formulations differ in the second factor. In Fermi's original formulations this factor was the square of (Nordheim and Yost, 1937)

$$T_\xi = (\psi_f, \tfrac{1}{2}(\tau_{\xi 1} + \tau_{\xi 2} + \ldots + \tau_{\xi n})\psi_i), \qquad \ldots\ldots(17)$$

while in the formulation of Gamow and Teller (1936) it is the sum of the squares of three matrix elements,

$$|Y_{\xi x}|^2 + |Y_{\xi y}|^2 + |Y_{\xi z}|^2, \qquad \ldots\ldots(18)$$

where

$$Y_{\xi x} = (\psi_f, \tfrac{1}{2}(\tau_{\xi 1}s_{x1} + \tau_{\xi 2}s_{x2} + \ldots + \tau_{\xi n}s_{xn})\psi_i) \qquad \ldots\ldots(18a)$$

and $Y_{\xi y}$ and $Y_{\xi z}$ are defined in a similar way. Finally, it is possible that the second factor is a sum of (18) and the square of (17) multiplied with a numerical factor. This would be the third formulation, a particular case of which has been discussed recently by Critchfield and Wigner (1941). It may be worth while to emphasize, at this juncture, the advantage of using the isotopic spins as variables in the wave function. Without doing so, it would be impossible to express the second factor for the β-transition probability in terms of matrix elements, since τ_ξ transforms a proton into a neutron and vice versa. Hence, if the matrix element is to be finite, the number of protons must be either greater by one or smaller by one in ψ_f than it is in ψ_i.

Before describing the consequences of the three different formulations on the basis of the 1st and 2nd approximations, it may be well to make a more general remark on the magnitude of matrix elements to be expected in nuclei.

It is in most cases relatively easy to estimate by closure the sum of the squares of the matrix elements between one state and all other states. Thus, for example,

$$\sum_f |(\psi_f, \tfrac{1}{2}(\tau_{\xi 1} + \ldots + \tau_{\xi n})\psi_i)|^2 = (\psi_i, \tfrac{1}{4}(\tau_{\xi 1} + \ldots + \tau_{\xi n})^2\psi_i). \qquad \ldots\ldots(19)$$

The last quantity has the value $\frac{1}{4}$ if ψ_i belongs to a $(\frac{1}{2}\ \frac{1}{2} + \frac{1}{2})$ supermultiplet. Furthermore, because of the breakdown of the Hartree model, most operators have matrix elements between almost any pair of states. If this were true for $\frac{1}{2}(\tau_{\xi 1} + \ldots + \tau_{\xi n})$, the matrix element for any single ψ_f would be expected to be quite small since the sum of the squares of so many of them is only $\frac{1}{4}$. We

can expect, therefore, sizable matrix elements only if some selection rule restricts ψ_f to a small group of states (members of one single supermultiplet or some similar group). The breakdown of the Hartree model is essential for this conclusion because the Hartree model would limit ψ_f to the relatively small number of states which can be reached from ψ_i by a single particle jump.

We now take up the discussion of the three alternative formulations (Wigner, 1939). The matrix element (17) plays the same rôle in our 2nd approximation as the matrix element of S_x in ordinary spectroscopy. It is finite only between two wave functions of the same T-multiplet, and its value between the $T_\zeta - 1$ and T_ζ members of a T-multiplet is

$$T_\xi = \tfrac{1}{2}(T - T_\zeta + 1)^{\frac{1}{2}}(T + T_\zeta)^{\frac{1}{2}}. \qquad \ldots\ldots(20)$$

Hence, in the first formulation of Fermi's theory, only transitions between members of the same T-multiplet are allowed. Among the transitions between the normal states of two elements, only those should be " allowed " which take place between the two members of a $T = \frac{1}{2}$ T-multiplet (cf., e.g., figure 6), and all other transition probabilities should be much smaller.

In the second or Gamow-Teller formulation the matrix elements responsible for the transition are of the form (18*a*). In addition to τ they contain also σ, so that they cannot be restricted any more to a small number of final states on the basis of the 2nd approximation. They can be calculated, however, using the 1st approximation and, since the operator does not contain the space co-ordinates, it will give finite matrix elements only between states of the same supermultiplet. This is less rigid than the restriction imposed by the first formulation to transitions within one T-multiplet, but the number of allowed transitions is still quite small. According to this formulation, the transitions between the normal states of $n = 4k + 2$ elements with $T_\zeta = -1, 0$ and $+1$ should be allowed because they are all members of a (1 0 0) super multiplet (figure 5). In addition to this, of course, the transitions between the members of the $(\frac{1}{2}\,\frac{1}{2} \pm \frac{1}{2})$ multiplet remain allowed. It is well to remember, however, that in the Gamow-Teller formulation, if our general remarks about the magnitude of matrix elements apply, all transition probabilities should be small, and rather erratic in their magnitude, unless the supermultiplets exist, i.e., unless the 1st approximation gives reasonably good wave functions.

The calculation of (18) for both series ^{6}He–^{6}Li, ^{10}C–^{10}B, etc., and ^{3}H–^{3}He, ^{7}Be–^{7}Li, etc., can be carried out on the basis of the 1st approximation (Wigner, 1939; Grönblom, 1939), but will not be given here. Since the $Y_{\xi z}$, etc., are components of a vector operator in the spin, the matrix element for singlet-singlet transitions will vanish and, similarly, between two states with $J = 0$. These transitions were allowed (although they could hardly be observed) in the first formulation, and are again allowed in the third formulation. The only other major difference between the second and third formulations is that in the third formulation the matrix elements for the ^{3}H–^{3}He, ^{7}Be–^{7}Li, etc. series can be expected to be large even if only the 2nd approximation is good

In the third formulation, therefore, the essential equality of the transition probabilities of the two series is coupled with the validity of the 1st approximation.

We now turn to the experimental material. If we go through the development, illustrated in figure 4, for a $(\frac{1}{2}\ \frac{1}{2} \pm \frac{1}{2})$ supermultiplet we see that it gives, unless $L=0$, two T-multiplets, one with $J=L-\frac{1}{2}$ and another with $J=L+\frac{1}{2}$. One of these is the lowest state for both $T_\zeta=-\frac{1}{2}$ and $\frac{1}{2}$. In the first formulation, from the normal state of the element $T_\zeta=-\frac{1}{2}$ only the transitions to the normal state of the element $T_\zeta=\frac{1}{2}$ are allowed. These transitions are, therefore, not expected to yield any γ rays. In the Gamow-Teller and the third formulations, however, transitions to all states of the same supermultiplet are allowed, i.e. to both the $J=L-\frac{1}{2}$ and the $J=L+\frac{1}{2}$ state of the element $T_\zeta=\frac{1}{2}$. Thus, in the Gamow-Teller and the third formulation, the transition from $T_\zeta=-\frac{1}{2}$ to $T_\zeta=\frac{1}{2}$ will be expected to go rather frequently to an excited state and to yield γ rays (Grönblom, 1939). Experimentally, γ rays were observed to be associated with many of these transitions, which speaks strongly in favour of the Gamow-Teller or the third formulation.

Konopinski (1942) gives a detailed survey of the experimental material for the lifetime of β-active substances. He finds that *ft*, which is g^{-2} times the reciprocal square of the matrix element, ranges between 1700 and 6200 for the 16 known β transitions* from a $T_\zeta=-\frac{1}{2}$ to a $T_\zeta=\frac{1}{2}$ nucleus. The *ft* are not constant within experimental error, which also speaks against the first formulation. According to the second and third formulations, there should be fluctuations in this quantity because the magnitude of (18) depends on L and J. There are indications that the fluctuations have the right direction, but the evidence, in its present form, is inconclusive. The fact that *ft* is roughly constant up to the ^{41}Sc–^{41}Ca transition must be considered as strong evidence, if the Gamow-Teller formulation is correct, that the 1st approximation is reasonable for the normal states of these elements. If the third (or first) formulation is correct, it can be considered as supporting only the 2nd approximation since the magnitude of (17) depends only on this approximation.

Four examples (^{6}He–^{6}Li, ^{10}C–^{10}B, ^{18}F–^{18}O, ^{26}Al–^{26}Mg) of a transition within a (1 0 0) multiplet give small values (1000 to 6000) for *ft*. This supports the the second or third formulation, *and the 1st approximation*. In these cases initial and final states of the transition cannot belong to the same T-multiplet (cf. figures 3 and 5), and the validity of the 2nd approximation alone could not explain the magnitude of the matrix elements. However, five other transitions (^{10}Be–^{10}B, ^{14}C–^{14}N, ^{22}Na–^{22}Ne, ^{30}P–^{30}Si, ^{34}Cl–^{34}S) of a similar nature have *ft* values which are a hundred times too large. In spite of numerous attempts to explain this, it has remained one of the most puzzling phenomena (Cooper and Nelson, 1941; Oppenheimer, 1941).

In addition to the above transitions there is only one, ^{12}B–^{12}C, which has a small *ft* value, namely 6600. There is no explanation known for the magnitude

* The energy relations for these transitions are given in table 1.

of the corresponding matrix element. It is satisfactory, however, that this is the only exception and that all other transitions, which should be forbidden if the 1st approximation is valid, yield *ft* values which are at least 30 times larger than those of the allowed transitions.

§ 13. GENERAL DISCUSSION OF CALCULATIONS OF MASS DEFECTS

The importance of mass defects derives mainly from the fact that they determine the stability of nuclei. Thus, for example, a nucleus is stable against β disintegration if its mass is lower (its mass defect higher) than that of either adjoining isobar (cf. Goeppert-Mayer, 1935). Otherwise it is unstable, and only in a few exceptional cases is the lifetime of a nucleus so long that it can be considered to be stable for practical purposes, even though its mass is slightly higher than that of one of the adjoining isobars (Weizsäcker, 1936 ; Sachs, 1940).

The importance of mass defects and the accuracy with which they can be determined explain the great number of attempts which have been made to correlate them. We cannot aim at giving an even approximately complete review of all these calculations, and shall restrict ourselves to those which bear most closely on the validity of our 1st approximation. For this reason we shall completely omit the extensive literature on the deuteron, since in this case the parity defines the partition quantum numbers, and these necessarily become good quantum numbers.

Generally speaking, the calculations on mass defects can be divided into two groups. The calculations of the first group use a definite model, such as the single-particle or alpha-particle model. The calculations of the second group endeavour to arrive at an estimate of the mass defect on the basis of somewhat crude but more general assumptions. The calculations of the first group give a much more detailed picture of the nucleus than those of the second group and yield results concerning not only mass defects but also other properties of the nucleus, such as spin and magnetic moment. Naturally, all the results which can be derived on the basis of general considerations can be derived, and sometimes more easily derived, on the basis of a special model.

We shall neglect henceforth the spin-dependent part of the Hamiltonian (the H_1 of § 9) because it seems to play an important rôle only in odd-odd nuclei. Our Hamiltonian will contain, apart from the electrostatic interaction, only ordinary and Majorana forces, defined in (6). This constitutes an additional specialization of the interaction to a two-particle interaction (Janossy, 1939; Primakoff and Holstein, 1939). The considerations of the previous sections were independent of this assumption, and also of the assumption necessary for the succeeding sections, that the $V_M(r)$ in (6) has roughly the form of a reversed bell.

It is, perhaps, well to emphasize that the following developments lag, from the point of view of mathematical rigour, far behind the developments of the

previous sections. This is manifested by the need of introducing either a model or rather crude assumptions. Both amount to *assuming* a form or a general shape of the wave function which in reality is *determined* by the Hamiltonian.

§ 14. THE POTENTIAL-WELL MODEL

In this model (Heisenberg, 1935; Majorana, 1933; Bartlett, 1932 a, 32 b; Elsasser, 1933-35; Guggenheimer, 1934) the wave function is calculated by assuming a Hamiltonian in which the interaction between the particles is replaced by a common potential. This potential has the value $-D$ inside of a sphere of radius R', which is called the potential well. The value of the potential is zero outside the well.

Evidently the model can be considered to be self-consistent only if R' is very nearly equal to, though perhaps somewhat larger than, the nuclear radius R, at which the density of the particles has dropped to one-half of its value at the centre. It would be very desirable if this condition were checked explicitly.

The potential-well model is evidently a single-particle model in which the total wave function is a product of single-particle wave functions. The mathematical problem of obtaining the latter is fairly simple (Margenau, 1934). It turns out that the lowest single-particle state is an s state, the next one a p state. This is succeeded by a d and an s state which lie rather close to each other, followed by f and p states. Pauli's principle permits only two protons and two neutrons to be accommodated in the lowest s shell, six protons and six neutrons in the succeeding p shell. The first shell is completed at ^{4}He, the second at ^{16}O, and the next, s–d shell, should be complete at ^{40}Ca. Although a more careful study of the nuclear mass defects does reveal discontinuities at these points (Bethe and Bacher, 1936, § 34),* indicating the completion of shells, these discontinuities are many times smaller than would follow from the potential-well model, which predicts discontinuities as large as 21 mMu. at ^{16}O and 8·5 mMu. at ^{40}Ca.

Predicting too large discontinuities in the mass-defect curve at the end of shells is a common failure of all single-particle models. The short-range correlations between the positions of the particles can be described only by the superposition of several configurations and are, therefore, not adequately given by any single-particle model. For this reason, a wave function containing a single configuration is not adequate for the calculation of the energy of nuclear forces which have short range. However, it can be used for the calculation of the electrostatic and kinetic energies.

The results of the calculations are most simple if n is very large (statistical model). In this case one can replace R' by R and the addition of the energies of the individual particles by integration. Neither the kinetic energy KE nor the electrostatic energy EE depend in this case on the depth of the well, and we have

$$\mathrm{KE} = (9/20)(6\pi^2)^{1/3}(\hbar^2/MR'^2)(\Lambda_1^{5/3} + \Lambda_2^{5/3} + \Lambda_3^{5/3} + \Lambda_4^{5/3}). \quad \ldots\ldots (21)$$

* Wigner (1937 b) obtains, for the actual magnitude of the discontinuity, 5 mMu. and 1·5 mMu. at $n = 16$ and $n = 40$, respectively. Cf. also Elsasser (1937).

Here, $\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4$ denote the number of particles in the four spin states ($\tau=\sigma=1$, $\tau=-\sigma=1$, $-\tau=\sigma=1$, $\tau=\sigma=-1$). It follows easily from the general theory that the Λ can be interpreted as partition quantum numbers in the sense of § 7, and (8) and (9) permit us to express them by means of n, P, P', P'' The result is, if P, P' and P'' are small compared to n (Wigner, 1940),

$$\text{KE}=14{\cdot}6n\,[1+(20/9n^2)(P^2+P'^2+P''^2)+\ldots]\ \text{mMu}. \qquad \ldots\ldots(22)$$

Herein, R' has been replaced by R which is taken from (5).

The electrostatic energy becomes

$$\text{EE}=\frac{3n_P(n_P-1)e^2}{5R}(1-0{\cdot}76n_P^{-2/3}). \qquad \ldots\ldots(23)$$

The first, and larger, of the two terms of (23) is the average electrostatic energy of n_P protons distributed at random in a sphere of radius R. This term has been used for the total electrostatic energy in (3). The second, much smaller, term of (23) is the electrostatic exchange energy. The Pauli principle prevents two protons with parallel spins from occupying the same position, and even a very close approach between them becomes improbable. This tends to decrease the electrostatic energy and gives rise to the second term of (23). On the other hand, two protons with antiparallel spin are more likely to be close to each other than would correspond to a uniform distribution. The free-particle model fails to describe this effect, which would give a positive contribution to the electrostatic energy and would tend to cancel the second term of (23). This may be, therefore, considerably in error (Wigner, 1940; Present, 1941) and has been omitted in (3). This discussion shows that the free-particle model is sufficiently accurate for the calculation of the electrostatic energy because the uncertainties, due to short-range correlations, are only of the order of the second, rather insignificant, term of (23).

The expression for the kinetic energy is not only more complicated for smaller n than it is for very large n, but also depends on the depth D of the potential well. If D were infinitely large, the wave function of every particle would vanish on the boundary of the well. Hence, the distance R at which the density of the particles drops to one-half of its value at the centre would be smaller than R' by a magnitude of the order of the wave-length of the particles. Thus the effective volume occupied by the particles would be smaller than the volume of the well (Slater, 1931; Wick, 1934) and their kinetic energy greater than (22). There would be a "surface energy" of the amount (Feenberg, 1941 b; Volkoff, 1942) $28n^{2/3}$ in addition to the volume energy (22).

If D is finite, the wave functions will vanish only asymptotically outside the well and the effective volume occupied by the particles will be larger, the surface energy smaller, than for infinite D (Slater, 1931; Wick, 1934; Feenberg, 1941 b; Volkoff, 1942). The most reasonable way to estimate the depth of the well is to postulate that the binding energy of the least strongly bound particle shall be equal to the experimental value for this quantity. Since the kinetic energy of

the least strongly bound particle is 5/3 times the average kinetic energy, this leads to $D=(5/3)14{\cdot}6+8{\cdot}5=33$ mMu. (Feenberg, 1941 b; Volkoff, 1942). The surface kinetic energy calculated with this D is much smaller than is the actual value,* indicating that the potential energy gives a major contribution to this quantity (cf. § 17).

Equation (22), or equivalent equations, are commonly used for estimating the kinetic energy of the nuclear particles. It may be useful to remark, therefore, that (22) is only a lower limit for this quantity, representing, in the sense of the uncertainty principle, the minimum kinetic energy necessary to concentrate n particles within a sphere of radius (5). The kinetic energy, associated with the actual wave function is no doubt considerably greater, since this not only concentrates all the particles within a well but also gives increased probability for attracting particles to be close to, for repelling particles to be far from, any given particle (Landshoff, 1941).

§ 15. THE HARTREE MODEL

It has been emphasized in the preceding section that one cannot expect to arrive at a satisfactory description of nuclear binding energies on the basis of the single-particle model. Nevertheless, some progress can be achieved by calculating the energy associated with the wave functions obtained in the preceding section. These calculations are usually referred to as being based on the Hartree model, although, as has been pointed out by Breit, they are more closely related to Slater's treatment (Slater, 1929; Condon and Shortley, 1936) of complex spectra than to Hartree's ideas.

The normal state configuration of the nuclei between ^{4}He and ^{16}O contains four s orbits and a sufficient number of p orbits to complete the configuration. Thus, for example, the configuration for eleven particles (^{11}B, ^{11}C, etc.) is s^4p^7. Hence, in the free-particle model, one can associate a great many wave functions with the normal states of the nuclei, since one can choose, subject to the restrictions of the Pauli principle, any of the three p orbits for each of the seven particles in p states. The " correct linear combinations " of these wave functions can be obtained by solving the secular equation of the real Hamiltonian associated with these wave functions. This can be done either directly—which is a very arduous task—or by using the artifices given by Slater (Slater, 1929; Condon and Shortley, 1936). Even then the computational difficulties remain prohibitive except for the simplest cases (Bethe and Bacher, 1936, § 36). However, if one uses a Hamiltonian for which the 1st approximation applies, one can make use of the transformation properties of the wave functions with respect to permutations and write down the correct linear combinations almost at once (Feenberg and Phillips, 1937). The wave functions obtained have definite partition and azimuthal quantum numbers. They are not simple products of single-particle wave functions, but sums of several such products. Therefore, the positions of

* Weizsäcker (1935), and Bethe and Bacher (1936) find that the observed binding energies can best be fitted with a surface energy of $14{\cdot}3\ n^{2/3}$ mMu.

the particles are no longer statistically independent but show correlations of the kind postulated in the preceding section. However, it must be conceded that these correlations are not as intimate as they would be for an accurate solution of the eigenvalue problem.

The potential energy within the p shell is the sum of the energies arising from ordinary and Majorana forces. The former can be expressed, just as in atomic spectroscopy (Slater, 1929; Condon and Shortley, 1936), in the terms of the two integrals

$$\begin{aligned} L_0 &= \int \psi_m(r)^* \psi_m(r')^* V_0(|r-r'|) \psi_m(r) \psi_m(r') dr\, dr', \\ K_0 &= \int \psi_m(r)^* \psi_{m'}(r')^* V_0(|r-r'|) \psi_{m'}(r) \psi_m(r') dr\, dr'. \end{aligned} \quad \ldots\ldots(24)$$

Herein $V_0(r)$ is the potential of the ordinary forces, ψ_m and $\psi_{m'}$ are two different wave functions for the p state. The former, larger integral is often called the direct, the second the exchange integral. One can easily prove (Slater, 1929; Condon and Shortley, 1936) that the coefficient of the former is $\frac{1}{2}n(n-1)$, where n is the number of particles in p states. The expressions for the potential energy of the ordinary forces for the different states of a configuration differ only in the coefficient of K_0. Thus the ordinary forces give only small energy differences between the different terms of a configuration.

The situation is radically different for the energy due to Majorana forces. This energy can also be expressed in terms of two integrals,

$$\begin{aligned} L_M &= \int \psi_m(r)^* \psi_m(r')^* V_M(|r-r'|) \psi_m(r) \psi_m(r') dr\, dr', \\ K_M &= \int \psi_m(r)^* \psi_{m'}(r')^* V_M(|r-r'|) \psi_{m'}(r) \psi_m(r') dr\, dr', \end{aligned} \quad \ldots\ldots(25)$$

(V_M gives the dependence of the Majorana interaction on distance, cf. (6)), but the coefficient of L_M is no longer the same for the different states of a configuration. Since it is, furthermore, very probable† that V_M is about four times larger than V_0, the Majorana forces will cause extremely large energy differences between the terms of a configuration. This is quite reasonable because the Majorana forces can give both attraction and repulsion. They give attraction between those particles in the co-ordinates of which the wave function is symmetric; they give repulsion between those particles in the co-ordinates of which the wave function is antisymmetric. Thus even the sign of the Majorana energy will be different for terms with different symmetry characters, even if they arise from the same configuration. It is therefore likely (Wigner, 1937 b) that terms with the same symmetry character arising from different configurations are more closely grouped together than are the terms of various symmetry characters within a configuration. In this respect the situation is opposite to that prevailing in atomic spectra. This circumstance has important consequenses for the density of levels. Because of the dependence of energy

† This follows from the so-called saturation arguments, summarized, for instance, by Breit and Wigner (1938).

on symmetry character there are characteristic oscillations in the level density as a function of T (Bardeen 1937; Bardeen and Feenberg, 1938; Motz and Feenberg, 1938).

Reasonable assumptions concerning ψ_m, V_0 and V_M give values of the direct integrals L which are 10 to 12 times larger than the corresponding exchange integrals K. Summarizing, we find that the total potential energy is, in the present picture, a linear combination of the integrals L_0, K_0, L_M, K_M. K_0 is so small that the corresponding term can be safely disregarded; the coefficient of L_0 is always the same, so that it does not influence the relative position of terms. The coefficient of L_M is the most important, and we shall determine it presently. The order of the terms for which L_M has the same coefficient will be determined by the coefficient of K_M.

Since the correct linear combinations of the various s^4p^n wave functions are determined by the symmetry properties of the system (i.e., by the quantum numbers (P, P', P'') and the azimuthal quantum number), the coefficients of the L_0, K_0, L_M, K_M are independent of the shape $V_0(r)$, $V_M(r)$ of the interaction. This enables one to find some of these coefficients very easily by using a $V_0(r)$ or $V_M(r)$, for which the potential energy can be calculated directly. This is the case for a long-range interaction, i.e. a $V_0(|\mathbf{r}-\mathbf{r}'|)=V_0$ or $V_M(|\mathbf{r}-\mathbf{r}'|)=V_M$, which has the same value, independent of the relative position of the particles within the nucleus. The direct calculation of the energy will be carried out in § 16. It shows that the potential energy of ordinary forces between the n particles in p states is $\frac{1}{2}n(n-1)V_0$. The potential energy of Majorana forces is also independent of the azimuthal quantum number of the term in question, but does depend on its partition quantum numbers. It is

$$\Xi V_M=[-\tfrac{1}{8}n^2+2n-\tfrac{1}{4}(P^2+4P+P'^2+2P'+P''^2)]V_M. \quad \ldots\ldots(26)$$

On the other hand, the exchange integrals K become zero for very long-range forces, while the direct integrals L assume the values V_0 and V_M, respectively. Comparison with the above values for the potential energy shows, therefore, that in the linear combination of L_0, K_0, L_M, K_M, which gives the potential energy in the Hartree picture, the coefficient of L_0 must be $\frac{1}{2}n(n-1)$, that of L_M must be Ξ. The determination of the coefficient of K_M requires (Feenberg and Phillips, 1937) more explicit calculations.

Since L_M is negative ($V_M(r)$ is negative), the general position of a term will be lower the lower its partition quantum numbers are. This is opposite to the corresponding rule in spectroscopy, but in agreement with the considerations of § 8, in particular with equation (11). The preceding consideration shows at the same time that the general order of the terms in the Hartree model agrees with their general order in the long-range force model. This holds, not only for our p shell, but quite generally.

The first two columns of the following tables (Feenberg and Phillips, 1937) give the combinations of the partition and azimuthal quantum numbers which the configuration in question yields. (They substitute for the combinations of

spin and azimuthal quantum numbers which are determined in ordinary spectroscopy by Hund's rules.) The last four columns give the coefficients of the four integrals (24), (25) for the energy of this term.

Table 5

$n=7$ (mass 11)						$n=8$ (mass 12)					
$(P\,P'\,P'')$	L	L_0	K_0	L_M	K_M	$(P\,P'\,P'')$	L	L_0	K_0	L_M	K_M
$(\frac{1}{2}\,\frac{1}{2}\,-\frac{1}{2})$	1	21	−24	6	21	(0 0 0)	0	28	−32	8	28
,,	2	,,	−26	,,	19	,,	2	,,	−35	,,	15
,,	3	,,	−29	,,	16	,,	4	,,	−42	,,	18
,,	4	,,	−33	,,	12	(1 1 0)	1	,,	−41	4	31
$(\frac{3}{2}\,\frac{1}{2}\,\frac{1}{2})$	1	,,	−30	3	24	,,	2	,,	−43	,,	29
,,	2	,,	−32	,,	22	,,	3	,,	−46	,,	26
,,	3	,,	−35	,,	19						

As was mentioned before, the relative position of the terms with the same partition quantum number is determined in this model by the coefficient of K_L. Since K_M is negative, the order of the terms follows in every case the order of their azimuthal quantum numbers L. This behaviour is again opposite to that encountered in atomic spectroscopy. It follows that $L=0$ for the normal state of the even elements and $L=1$ for the odd elements. This is in apparent agreement with the observed J values: $J=0$ for the even elements ($S=0$) and $J=1/2$ or 3/2 for odd elements ($S=1/2$). This applies not only for $n\leqslant 16$ but also for higher n (Hund, 1937 b) if one generalizes the second rule, for odd n, to state that L is not larger than the l of the shell which is being built up. We must not forget, however, that the magnetic moments show that L is in most cases not a very good quantum number.

Although the more general theory, which will be developed in §17, also yields $S=0$ for the normal states of even and $S=1/2$ for the normal states of odd elements, it fails to make any prediction for the azimuthal quantum numbers of the normal states. The fact that the L values predicted by the Hartree model are supported by experience indicates that it is, probably, not quite as far from reality as has been supposed for some time. The absence of discontinuities in the mass-defect curve of a magnitude predicted by the single-particle model (§14) shows that the actual normal state is a superposition of wave functions belonging to several configurations. This may be also responsible for the fact that the calculated binding energies fall far short of the empirical values. However, the lowest configuration, though not the only one, may play a sufficiently large rôle, so as to determine the J value of the normal state. It is evident that the effects of the shell structure will be less noticeable the less preponderant the lowest configuration is in the actual wave function. In the next section we shall trace the effect of the preponderance of the lowest Hartree wave functions on the electrostatic energy.

§ 16. ELECTROSTATIC ENERGY IN THE HARTREE MODEL

The simplicity of the application of the Hartree picture to the nuclei from He to O is quite surprising. However, the extension of these calculations beyond $n=16$, as well as their improvement by the inclusion of higher configurations, meets, except for $n=6$, substantial computational difficulties.* Thus only the wave functions of the elements from $n=4$ to $n=16$, in the form given by the unmodified Hartree picture, are easily available for the comparison mentioned in the last section.

The differences between the observed electrostatic energies and the values calculated on the basis of the " uniform " picture for the masses 7, 9, 11, 13, 15 are (cf. table 1) 0·26, 0·48, −0·05, 0·05, −0·12 mMu. Although some of these differences are probably within experimental error, it does appear that they show a decreasing tendency, superimposed upon which is an oscillation.

Both of these effects are in the direction predicted by the single-particle model. According to this, the average radius of the orbit of the last proton should remain nearly constant over the whole shell. Thus the uniform picture can be expected to give too small radii for small n, too large radii for large n, i.e., too much electrostatic energy for small, too little for large n. The above figures indicate that the actual radii are somewhere between the values given by the uniform and Hartree pictures.

There are two effects which can be made responsible for the periodic fluctuations in the electrostatic energy. The first (Feenberg and Phillips, 1941; Brown and Inglis, 1939) is given immediately by the oscillations of the coefficients of the electrostatic exchange integral in the Hartree model. The coefficients are $2/3$, $-16/3$, $-14/3$, $-32/3$, $-30/3$ for the masses 7, 9, 11, 13, 15 respectively. Although the calculated value of the electrostatic exchange integral is very small (0·03 mMu.), the oscillation of the above factors explains a good part of the oscillation of the electrostatic energies. The second effect (Bethe, 1938 b; Brown and Inglis, 1939) is caused by the expansion of the orbits, due to the electrostatic repulsion itself. This, calculated in a straightforward fashion by means of a single-particle model, would give an even stronger fluctuation than is observed (Bethe, 1938 b; Brown and Inglis, 1939). Evidently the actual expansion of the orbits is much smaller than the single-particle model would indicate.

§ 17. THE " UNIFORM " MODEL

The " uniform " model does not assume any particular mathematical form for the wave function, but attempts to use only assumptions of a general nature concerning the relative positions, etc., of particles. Chief among these assumptions is that of the existence of a rather uniform nuclear matter which has much

* Calculations aimed at an improvement of the Hartree model were carried out by numerous authors: cf. Inglis, 1937, 38 b; Kroeger, 1938; Margenau and Carroll, 1938; Margenau, 1939; Tyrell, 1939; Carroll, 1940. The work of Horvay (1939) is based on the variational method, and gives, therefore, a lower limit for the binding energy: cf. further, Fuchs, 1940; Spain, 1940.

the same properties in different parts of the nucleus and does not change much from one nucleus to another. This assumption is also incorporated in Bohr's liquid-drop model (Bohr, 1936; Bethe, 1937), which, however, is often regarded as going further in attributing to the nuclear matter specific properties of ordinary liquids. The property of uniformity is common to solids, liquids, degenerate and ordinary gases, plastics and, probably, to many other forms of matter, among which is, we hope, the " nuclear material ".

We intend to give an estimate of the energy of the Majorana interaction, but must begin with a somewhat more general mathematical discussion. Let us denote the operator which interchanges the space co-ordinates of the particles i and k by P_{ik}. We have then the identity

$$P_{ik}\psi = -\tfrac{1}{4}[1+(\boldsymbol{s}_i \cdot \boldsymbol{s}_k)]\,[1+(\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_k)]\psi. \qquad \ldots\ldots(27)$$

One can ask for the probability p such that, for example, P_{12} be 1 (or that it be -1). For this purpose we can decompose the wave function into a symmetric and an antisymmetric part,

$$\psi_s = \tfrac{1}{2}(1+P_{12})\psi; \qquad \psi_a = \tfrac{1}{2}(1-P_{12})\psi. \qquad \ldots\ldots(28)$$

This gives, since the square of P_{12} is unity,

$$P_{12}\psi_s = \psi_s; \qquad P_{12}\psi_a = -\psi_a, \qquad \ldots\ldots(28a)$$

and the probability p_s becomes

$$\left.\begin{aligned} p_s &= (\psi_s, \psi_s) = \tfrac{1}{4}[(1+P_{12})\psi, (1+P_{12})\psi] = \tfrac{1}{4}[\psi, (1+P_{12})^2\psi] \\ &= \tfrac{1}{2}[\psi, (1+P_{12})\psi] = \tfrac{1}{2} + \tfrac{1}{2}(\psi, P_{12}\psi). \end{aligned}\right\} \qquad \ldots\ldots(29)$$

Since the probability that P_{12} be 1 is the same as for any other pair i, k, one can write also

$$\tfrac{1}{2}n(n-1)p_s = \tfrac{1}{4}n(n-1) + \tfrac{1}{2}\sum_{i\neq k}{}' \tfrac{1}{2}[\psi, -\tfrac{1}{4}(1+(\boldsymbol{s}_i \cdot \boldsymbol{s}_k))(1+(\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_k))\psi]. \qquad \ldots\ldots(30)$$

Every wave function ψ with definite partition quantum numbers (P P' P'') is a characteristic function of $\Sigma'\frac{1}{4}(1+(\boldsymbol{s}_i \cdot \boldsymbol{s}_k))(1+(\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_k))$, just as, in ordinary spectroscopy, every wave function which belongs to a definite multiplet system S is a characteristic function of $\Sigma'\frac{1}{2}(1+(\boldsymbol{s}_i \cdot \boldsymbol{s}_k))$. The characteristic value of the former can be calculated just as Dirac (1935, § 61) calculated the characteristic value of $\Sigma'\frac{1}{2}(1+(\boldsymbol{s}_i \cdot \boldsymbol{s}_k))$, although, of course, one can use several other methods of calculation as well. The result is (Wigner, 1937 b)

$$\left.\begin{aligned} &\tfrac{1}{2}\Sigma' - \tfrac{1}{4}(1+(\boldsymbol{s} \cdot \boldsymbol{s}_k))(1+(\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_k))\psi = \Xi\psi, \\ &\Xi = -\tfrac{1}{8}n^2 + 2n - \tfrac{1}{2}(P^2 + P'^2 + P''^2 + 4P + 2P'). \end{aligned}\right\} \qquad \ldots\ldots(31)$$

This gives, with (30), for the probability that P_{12} be 1

$$p_s = \tfrac{1}{2} + \frac{1}{n(n-1)}\Xi; \qquad \ldots\ldots(31a)$$

the probability that P_{12} be -1 becomes

$$p_a = \tfrac{1}{2} - \frac{1}{n(n-1)}\Xi. \qquad \ldots\ldots(31b)$$

We can now ask for the probability $g_s(r_{12})$ of a certain distance between the particles 1 and 2, *if* P_{12} is 1. This is given by the equation

$$p_s g_s(r_{12}) = \sum_{\sigma,\tau} \int |\psi_s|^2 dr_3 dr_4 \ldots dr_n. \qquad \ldots\ldots(32a)$$

Since ψ_s is a symmetric function of r_{12}, the g_s will start with zero slope at $r_{12}=0$. For larger r_{12} it will approach a constant value which is approximately the reciprocal volume of the nucleus, since the integral of $g_s(r_{12})$ is unity. Therefore, g_s will have some shape similar to that of the upper curve of figure 8.

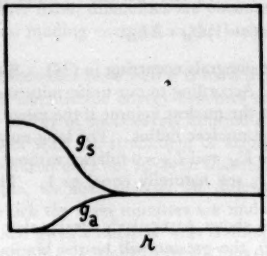


Figure 8.

Probability of a distance r between two nuclear particles. g_s refers to symmetrically, g_a to anti-symmetrically coupled particles.

Similarly, we can ask for the probability $g(r_{12})$ of a distance r_{12} between the particles 1 and 2 if P_{12} is -1. This is given by

$$p_a g_a(r_{12}) = \sum_{\sigma,\tau} \int |\psi_a|^2 dr_3 dr_4 \ldots dr_n. \qquad \ldots\ldots(32b)$$

However, since ψ_a is antisymmetric in r_{12}, it vanishes at $r_{12}=0$ and $g_a(r_{12})$ will have the general form of the lower curve of figure 8. The curves of figure 8 were drawn on the basis of the wave function of the statistical model (§ 14), and are correct, of course, only in a rough way.

We shall now make the assumption that the shapes of both g_s and g_a, whatever they are, are characteristic of the "nuclear matter", that they change very little from one nucleus to another, and that they are, apart from the surface, the same over the whole nucleus.

Under this assumption we can calculate at once the energy arising from Majorana or ordinary forces (Wigner, 1937 b). The energy of the Majorana interaction between the particles 1 and 2 is

$$\left.\begin{aligned}(\psi,\ P_{12}V_M(r_{12})\psi) &= (\psi_s+\psi_a,\ V_M(r_{12})(\psi_s-\psi_a)) \\ &= (\psi_s,\ V_M(r_{12})\psi_s) - (\psi_a,\ V_M(r_{12})\psi_a).\end{aligned}\right\} \qquad \ldots\ldots(33)$$

Since ψ_s and V_M are symmetric and ψ_a antisymmetric, the cross-products drop out. Furthermore, since

$$(\psi_s, V_M(r_{12})\psi_s) = \sum_{\sigma,\tau} \int |\psi_s|^2 V_M(r_{12}) d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_n,$$

$$= p_s \int g_s(r_{12}) V_M(r_{12}) d\mathbf{r}_1 d\mathbf{r}_2,$$

and since a similar expression holds for $(\psi_a, V_M(r_{12})\psi_a)$, we can write for the whole energy ME of Majorana forces, using (31*a*) and (31*b*),

$$\left.\begin{aligned} \mathrm{ME} &= \tfrac{1}{2}n(n-1)(p_s \int g_s V_M - p_a \int g_a V_M) \\ &= \tfrac{1}{2}n(n-1) \int \tfrac{1}{2}(g_s - g_a) V_M + \Xi \int \tfrac{1}{2}(g_s + g_a) V_M \\ &= \tfrac{1}{2}n(n-1) L'_M + \Xi L_M. \end{aligned}\right\} \quad \ldots\ldots(34)$$

The L'_M and L_M are the integrals occurring in (34). Since $V_M(r_{12})$ is negative, they are both negative. According to our uniformity assumption they are both inversely proportional to the nuclear volume if the range of the Majorana forces is small compared to the nuclear radius. For long-range forces, on the other hand, the fact that $L_M = V_M$ and $L'_M = 0$ follows without any assumption, since the integrals of g_s and g_a are naturally equal to 1. This gives equation (26) of § 15.

The first line of (34) shows particularly clearly that the larger g_a is in the neighbourhood of $r_{12} = 0$, the greater will be the binding energy. Therefore, the actual wave function will deviate from the wave function of the potential-well picture so as to make the maximum of g_a, and also the minimum of g_a, more pronounced. This modification will be limited, however, by the increase in kinetic energy which such a modification involves.

For the potential energy OE of ordinary forces a calculation analogous to that above yields

$$\mathrm{OE} = -\tfrac{1}{2}n(n-1)L_0 - \Xi L'_0, \quad \ldots\ldots(35)$$

where the definition of L_0 and L'_0 is analogous to that of L_M and L'_M.

The consequences of an expression for the total energy

$$E = \mathrm{KE} + \mathrm{EE} + \mathrm{ME} + \mathrm{OE} \quad \ldots\ldots(36)$$

(where KE, EE, ME and OE are defined in (22), (23), (34), (35)) have been investigated extensively (Wigner, 1937 b, 40; Barkas, 1939 *). A constant particle density corresponding to a nuclear radius very nearly equal to (5) has been used in all these calculations. This appears justifiable since the variation of particle density within the nucleus (Wigner, 1940 *; Feenberg, 1941 a), and from one nucleus to another (Present, 1941), have been found to be sufficiently

* The definition of L_s, L'_0, L_M, L'_M in this *Report* assigns to these quantities signs which are opposite to those used in Wigner, 1937, 40, and Barkas, 1939. This change in sign has been made in order to bring out more clearly the analogy between L_s and L_M of § 15 and § 17.

small to permit neglecting all variations. If the nuclear radius is given, only one additional constant, $L_M + L_0'$, enters into the comparison of the masses of isobars. The best value for that constant is about

$$L_M + L_0' = -(41/n)\ \text{mMu}. \qquad \ldots\ldots(37)$$

The agreement between (36) and the observed mass differences of isobars is satisfactory.

Subsequent experimental work has demonstrated the occurrence in nature of several nuclei, the stability of which follows from (36). The situation is similar for the binding energy differences of nuclei with not very different mass numbers. The calculations of Barkas (1939) proved in several cases to be more accurate than the measured values available at the time the calculations were made. On the other hand, difficulties are encountered (Wigner, 1940) in the interpretation of total binding energies on the basis of (36), particularly for heavier nuclei.

In spite of apparent successes for $n > 80$ it is hard to believe that (36) applies in this region, since the electrostatic energy becomes so very large. The continuous increase in the energy of successively emitted α particles in the natural radioactive series together with the attainment of stability at Pb shows clearly that the uniformity hypothesis does not apply in that region.

As a specific example, we shall discuss the differences between the mass defects of isobars for which the mass numbers are multiples of 4. In this case the lowest partition which gives a nucleus with isotopic spin T_ζ is $(P\,P'\,P'') = (T_\zeta\ 0\ 0)$ if T_ζ is even and $(P\,P'\,P'') = (T_\zeta\ 1\ 0)$ if T_ζ is odd (rules *a*, *b*, *c* of §7). This gives, by (22), for n times the KE of the normal states, the broken line of figure 9*b* (terms independent of T_ζ are omitted in all figures 9). Similarly, figure 9*a* gives n times the sum of ME and OE as functions of T_ζ. Figure 9*c* gives n times the electrostatic energy differences (4) for various n. In order to obtain the total mass differences between isobars, the three curves of figures 9*a*, 9*b*, 9*c* must be added. Those nuclei will be stable, the masses of which are, in the resulting diagram, lower than the masses of both adjoining isobars.

For low n, this will be the nucleus $T_\zeta = 0$, with equal number of protons and neutrons. For higher n, the points for larger T_ζ will become favoured, because of the increase in the electrostatic energy. However, as a result of the breaks in figure 9*b* the T_ζ will increase in jumps of two. Thus, for example $T_\zeta = 2$ lies lower than $T_\zeta = 1$ from $n = 36$ on, while $T_\zeta = 1$ will be lower than $T_\zeta = 0$ only from $n = 44$ on. Hence both $T_\zeta = 0$ and $T_\zeta = 2$ are stable for $n = 36$ and 40, while $T_\zeta = 1$ is unstable for all n. The same applies for all odd T_ζ, i.e. for all odd-odd nuclei of this series.

In the other series of even nuclei the even T_ζ are unstable, apart from $T_\zeta = 0$ (cf. table 3 or figure 5). Thus the only stable odd-odd nuclei should occur in the $4k+2$ series and have an equal number of protons and neutrons. This is very well borne out by Aston's well-known rule (Harkins, 1921, 31; Aston, 1933).

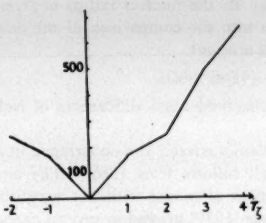


Figure 9a.

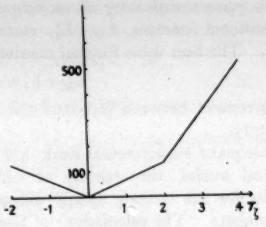


Figure 9b.

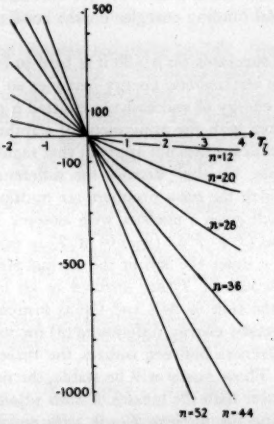


Figure 9c.

Figure 9. Calculation of the differences between the masses of isobars according to the " uniform " model. Figures 9*a* and 9*b* give nuclear potential and kinetic energies, multiplied by the number of particles. Figure 9*c* gives n times the electrostatic energy. The zero of the energy scale is so adjusted that all lines go through the origin of the co-ordinate system.

If one corrects the experimental mass differences between isobars (excepting those of table 3) for the electrostatic energy—as was done, for example, in table 4 of § 8—the remaining differences, containing only the KE, ME and OE terms, should be inversely proportional to n, the factor of $1/n$ being given by (22), (34), (35), (37). One can obtain a fair indication from table 4 to what extent this holds true.

§ 18. APPLICATION OF THE MOLECULAR VIEWPOINT TO THE THEORY OF NUCLEI

The considerations which will be sketched very briefly and incompletely in this last section do not belong, strictly speaking, to an article on the ***symmetry properties of nuclear levels***. It will deal with symmetry properties and also with nuclear levels, but the considerations which deal with symmetry properties originate in the theory of molecules rather than nuclei. They were developed several years ago by Dennison, Lennard-Jones, Mullikan, Placzek, Teller, Tisza, Van Vleck, Wilson and many others, and were applied to the problem of nuclei by Dennison, Inglis, Teller, Wefelmeier, Wheeler and many others. The most obvious difference between the considerations presented in the previous sections and those based on the molecular viewpoint is that the latter gives, so to say, a separate model for every nucleus—just as one has a separate model for every molecule. This makes the theory very interesting and gives it a beauty of detail which the previous sections lack—but it also makes the theory very difficult to review. It might be useful to mention at this point that the molecular viewpoint conflicts in no way with the general considerations up to § 15. It conflicts with the special models of § 15, § 16 and, perhaps, § 17, but all these models, together with the present one, conform to the general postulates of the sections preceding them (Hafstad and Teller, 1938). However, it is, in practice, possible to describe, by means of the molecular viewpoint, only those nuclear states which belong to the partitions (0 0 0) and $(\frac{1}{2}\,\frac{1}{2} \pm \frac{1}{2})$. Hence, the partition quantum numbers play no rôle in distinguishing between the states available for discussion by the molecular method. The latter is most useful in giving an estimate of the relative positions of terms which belong to the lowest partition, It is at this point that the models of the preceding sections fail most signally. the " uniform " model, in particular, permitting no conclusions to be drawn concerning the relative positions and azimuthal quantum numbers of these terms.

The molecular viewpoint owes its origin, however, to another flaw of the theories reviewed above. These, particularly the Hartree model, while describing the fluctuations of the binding energy in a relatively satisfactory way, fail to give a good account of the total binding energy. In order to overcome this difficulty, the whole nucleus is regarded as being built out of blocks which already contain a substantial part of the binding energy. The α particle is a suitable block of this nature, and the molecular viewpoint might be expected to be most useful for nuclei which can be built up out of α particles alone, such as ^{8}Be, ^{12}C, ^{16}O, and so forth. The other nuclei are regarded as containing, in addition to α particles, extra neutrons or protons which act as a " cement " similar to the " free " electrons in metals.

The total binding associated with the wave functions of the α-particle model comes, in fact, very much closer to the observed binding energies than that associated with the wave functions of the Hartree model. There is thus

little doubt that the former wave functions are less inaccurate than the latter. In spite of this, the writers must admit that they could not convince themselves that any strong support for the α-particle picture can be obtained from the detailed study (Wefelmeier, 1937; Hafstad and Teller, 1938) of mass defects. On the other hand, the discussion of the lowest states of the nuclei ^{8}Be, ^{12}C, ^{16}O is, at least in the case of ^{16}O (Dennison, 1940; Wheeler, 1937), definitely superior to the discussion which can be carried out by any of the previously described methods.

The ^{16}O nucleus consists, according to the α-particle picture, of four α particles, arranged at the corners of a regular tetrahedron. The rotation and vibration of this tetrahedron will give the excited states of the ^{16}O nucleus, and we shall now outline the quantum mechanical discussion of the rotational and vibrational states (cf., e.g., Dennison, 1931, or Teller, 1934).

Let us denote the wave function of the α particle containing the first four particles by $\alpha_{1\,2\,3\,4}$. It is, of course, a function of the space, ordinary and isotopic spin co-ordinates of the first four particles. It has the partition quantum numbers (0 0 0), which means, according to the general theory, that it is the product of a completely symmetric function of the space co-ordinates, multiplied by the antisymmetric function of the spin co-ordinates. It is usually assumed that (Breit and Feenberg, 1936)

$$\exp[-\gamma(r_{12}^2+r_{13}^2+r_{14}^2+r_{23}^2+r_{24}^2+r_{34}^2)] \qquad \ldots\ldots(38)$$

is a good approximation for the first factor of $\alpha_{1\,2\,3\,4}$ but this is not important for what follows. The second factor is a determinant, the main diagonal of which is

$$\delta(\sigma_1, 1)\delta(\tau_1, 1)\,.\,\delta(\sigma_2, 1)\delta(\tau_2, -1)\,.\,\delta(\sigma_3, -1)\delta(\tau_3, 1)\,.\,\delta(\sigma_4, -1)\delta(\tau_4, -1). \qquad \ldots\ldots(38\,a)$$

The wave functions of the ^{16}O nucleus, which are built up out of α particles have the form

$$\Sigma_p \epsilon_p\, g(\rho_{1\,2\,3\,4},\ \rho_{5\,6\,7\,8},\ \rho_{9\,10\,11\,12},\ \rho_{13\,14\,15\,16})\,\alpha_{1\,2\,3\,4}\,\alpha_{5\,6\,7\,8}\,\alpha_{9\,10\,11\,12}\,\alpha_{13\,14\,15\,16}, \qquad \ldots\ldots(39)$$

where $\rho_{1\,2\,3\,4}=\frac{1}{4}(r_1+r_2+r_3+r_4)$ stands for the co-ordinates of the centre of mass of the first α particle, etc. $\Sigma_p\epsilon_p$ means that the ensuing expression must be antisymmetrized (cf., e.g., Slater, 1930) to conform to the Pauli principle. The factor g will describe the relative motion of the α particles, i.e. their rotation about the common centre of mass and their vibration.

It is important for the following that, because of the antisymmetrization, g must be a symmetric function with respect to the interchange of the four ρ (Ehrenfest and Oppenheimer, 1931). Thus, for example, (39) would vanish identically if g were an antisymmetric function. The symmetry property of g plays a fundamental rôle in the spectra of molecules with tetrahedral symmetry (cf., e.g., Dennison, 1931, or Teller, 1934). In the present case it is usually referred to as the Bose statistics of the α particles. Its importance for nuclear

spectra has been pointed out by Wheeler (1937). Incidentally, (39) shows that the spin dependence of the wave functions in question is independent of the relative motion g of the α particles, i.e. independent of their state of rotation and vibration. (The same holds for their symmetry properties with respect to an interchange of their space co-ordinates.) Hence, they all belong to the same partition, which is obviously (0 0 0), since $T_\zeta = S_z = 0$ for all functions of the form (39).

From here on, the discussion centres around the symmetric function g and can be carried through exactly in the same way as in the theory of molecular spectra. A four-atomic molecule for which the equilibrium configuration is a regular tetrahedron has three different vibrational frequencies. We shall be interested only in those states in which not more than one vibrational quantum is present.

The wave function of a vibrationless state is a symmetric function of the variables which describe the distortion of the tetrahedron, and must be multiplied, therefore, by a symmetric function of the rotational co-ordinates. Such symmetric functions exist for $L = 0, 3, 4$, etc., not, however, for $L = 1, 2$. Hence the vibrationless state is compatible only with the angular momenta 0, 3, 4, etc. The important consequence of this phenomenon of " missing rotational states " is that the second rotational level of the vibrationless state lies very high because of its high L. Dennison estimates its position as between 4·4 and 5·5 mMu. from the normal state. It may, therefore, lie even higher than the state which has a pulsating vibration but no rotation. The relatively low position of this second rotationless state (Oppenheimer and Schwinger, 1939) caused difficulties for the one-particle picture.

REFERENCES

ALVAREZ, L. W. and BLOCH, F., 1940. *Phys. Rev.* **57**, 111.
ASTON, F. W., 1933. *Mass-Spectra and Isotopes* (London).
BARDEEN, J., 1937. *Phys. Rev.* **51**, 799.
BARDEEN, J. and FEENBERG, E., 1938. *Phys. Rev.* **54**, 809.
BARKAS, W. H., 1939. *Phys. Rev.* **55**, 691.
BARTLETT, J. H., 1932 a. *Phys. Rev.* **41**, 370.
BARTLETT, J. H., 1932 b. *Nature, Lond.*, **130**, 165.
BARTLETT, J. H., 1936. *Phys. Rev.* **49**, 102.
BETHE, H. A., 1937. *Rev. Mod. Phys.* **9**, 71.
BETHE, H. A., 1938 a. *Phys. Rev.* **53**, 842.
BETHE, H. A., 1938 b. *Phys. Rev.* **54**, 436.
BETHE, H. A., 1939. *Phys. Rev.* **55**, 1261.
BETHE, H. A., 1940. *Phys. Rev.* **57**, 260 and 390.
BETHE, H. A. and BACHER, R. F., 1936. *Rev. Mod. Phys.* **8**, 82.
BHABHA, H., 1938. *Proc. Roy. Soc.* A, **166**, 501.
BLOCH, F., 1940. *Phys. Rev.* **58**, 829.
BOHR, N., 1936. *Nature, Lond.*, **137**, 344.
BREIT, G., CONDON, E. U. and PRESENT, R. D., 1936. *Phys. Rev.* **50**, 825.
BREIT, G. and FEENBERG, E., 1936. *Phys. Rev.* **50**, 850.
BREIT, G. and STEHN, J. R., 1937. *Phys. Rev.* **52**, 396.

BREIT, G., THAXTON, H. M. and EISENBUD, L., 1939. *Phys. Rev.* **55**, 1018.
BREIT, G. and WIGNER, E., 1938. *Phys. Rev.* **53**, 998.
BROWN, H. and INGLIS, D. R., 1939. *Phys. Rev.* **55**, 1182.
BROWN, S. C., 1941. *Phys. Rev.* **59**, 954.
CARROLL, K. G., 1940. *Phys. Rev.* **57**, 791.
CASIMIR, H. B. G., 1940. *Physica*, **7**, 169.
CASSEN, B. and CONDON, E. U., 1936. *Phys. Rev.* **50**, 846.
CONDON, E. U. and SHORTLEY, G., 1936. *The Theory of Atomic Spectra* (Cambridge).
COOPER, E. P. and NELSON, E. C., 1941. *Phys. Rev.* **59**, 216.
CRITCHFIELD, C. and WIGNER, E., 1941. *Phys. Rev.* **60**, 412.
DELSASSO, L. A., WHITE, M. G., BARKAS, W. and CREUTZ, E. C., 1940. *Phys. Rev.* **58**, 586.
DENNISON, D. M., 1931. *Rev. Mod. Phys.* **3**, 280.
DENNISON, D. M., 1940. *Phys. Rev.* **57**, 454.
DIRAC, P. A. M., 1935. *The Principles of Quantum Mechanics* (Oxford).
EHRENFEST, P. and OPPENHEIMER, J. R., 1931. *Phys. Rev.* **37**, 333.
EISENBUD, L. and WIGNER, E., 1941. *Proc. Nat. Acad. Sci., Wash.*, **27**, 281.
ELSASSER, W. M., 1933. *J. Phys. Radium*, **4**, 549.
ELSASSER, W. M., 1934. *J. Phys. Radium*, **5**, 389 and 635.
ELSASSER, W. M., 1935. *J. Phys. Radium*, **6**, 473.
ELSASSER, W. M., 1936. *J. Phys. Radium*, **7**, 312.
ELSASSER, W. M., 1937. *Phys. Rev.* **51**, 55 and 144.
FARKAS, L. and SANDLER, L., 1939. *Trans. Faraday Soc.* **35**, 337.
FEENBERG, E., 1941 a. *Phys. Rev.* **59**, 149 and 593.
FEENBERG, E., 1941 b. *Phys. Rev.* **60**, 204.
FEENBERG, E. and KNIPP, J., 1935. *Phys. Rev.* **48**, 906.
FEENBERG, E. and PHILLIPS, M., 1937. *Phys. Rev.* **51**, 957.
FEENBERG, E. and PHILLIPS, M., 1941. *Phys. Rev.* **59**, 400.
FEENBERG, E. and SHARE, S., 1936. *Phys. Rev.* **50**, 488.
FEENBERG, E. and WIGNER, E., 1937. *Phys. Rev.* **51**, 95.
FERMI, E., 1934. *Z. Phys.* **88**, 161.
FOWLER, W. A., DELSASSO, L. A. and LAURITSEN, C. C., 1936. *Phys. Rev.* **49**, 561
FROHLICH, H., HEITLER, W. and KEMMER, N., 1938. *Proc. Roy. Soc.* A, **166**, 154.
FUCHS, K., 1940. *Proc. Roy. Soc.* A, **174**, 509.
GAMOW, G. and TELLER, E., 1936. *Phys. Rev.* **49**, 895.
GERJUOY, E. and SCHWINGER, J., 1941. *Phys. Rev.* **60**, 158.
GOEPPERT-MAYER, M., 1935. *Phys. Rev.* **48**, 512.
GRÖNBLOM, B. O., 1939. *Phys. Rev.* **56**, 508.
GUGGENHEIMER, K., 1934. *J. Phys. Radium*, **5**, 253 and 475.
HAFSTAD, L. R. and TELLER, E., 1938. *Phys. Rev.* **54**, 681.
HARDY, T. C., 1941. *Phys. Rev.* **60**, 167.
HARKINS, W. D., 1921. *Nature, Lond.*, **107**, 202.
HARKINS, W. D., 1931. *Phys. Rev.* **38**, 1270.
HAXBY, R. O., SHOUPP, W. E., STEPHENS, W. E. and WELLS, W. H., 1940. *Phys Rev.* **58**, 1035.
HEISENBERG, W., 1932 a. *Z. Phys.* **77**, 1.
HEISENBERG, W., 1932 b. *Z. Phys.* **78**, 156.
HEISENBERG, W., 1935. *Z. Phys.* **96**, 473.
HORVAY, G., 1939. *Phys. Rev.* **55**, 70.
HUND, F., 1937 a. *Z. Phys.* **105**, 202.
HUND, F., 1937 b. *Phys. Z.* **38**, 929.
INGLIS, D. R., 1937. *Phys. Rev.* **51**, 531.
INGLIS, D. R., 1938 a. *Phys. Rev.* **53**, 470.
INGLIS, D. R., 1938 b. *Phys. Rev.* **53**, 880.

INGLIS, D. R., 1939. *Phys. Rev.* **55**, 329.
JANOSSY, L., 1939. *Proc. Camb. Phil. Soc.* **35**, 616.
KELLOGG, J. M. B., RABI, I. I., RAMSEY, N. F. and ZACHARIAS, J. R., 1940. *Phys. Rev.* **57**, 677.
KEMMER, N., 1938. *Proc. Camb. Phil. Soc.* **34**, 354.
KONOPINSKI, E. J., 1942. *Beta Decay* ; to appear soon.
KROEGER, W. G., 1938. *Phys. Rev.* **54**, 1048.
LANDSHOFF, R., 1941. *Phys. Rev.* **59**, 906.
MAJORANA, E., 1933. *Z. Phys.* **82**, 137.
MARGENAU, H., 1934. *Phys. Rev.* **46**, 613.
MARGENAU, H., 1939. *Phys. Rev.* **55**, 1173.
MARGENAU, H. and CARROLL, K. G., 1938. *Phys. Rev.* **54**, 705.
MARGENAU, H. and WIGNER, E., 1939. *Phys. Rev.* **58**, 103.
MILLMAN, S. and KUSCH, P., 1941. *Phys. Rev.* **60**, 91.
MILLMAN, S., KUSCH, P. and RABI, I. I., 1939. *Phys. Rev.* **56**, 165.
MOTZ, L. and FEENBERG, E., 1938. *Phys. Rev.* **54**, 1055.
NORDHEIM, L. W. and YOST, F. L., 1937. *Phys. Rev.* **51**, 943.
NORDSIECK, A., 1940. *Phys. Rev.* **58**, 310.
O'NEAL, R. D. and GOLDHABER, M., 1940. *Phys. Rev.* **58**, 574.
OPPENHEIMER, J. R., 1941. *Phys. Rev.* **59**, 908.
OPPENHEIMER, J. R. and SCHWINGER, J., 1939. *Phys. Rev.* **56**, 1066.
PHILLIPS, M., 1940. *Phys. Rev.* **57**, 160.
PRESENT, R. D., 1941. *Phys. Rev.* **60**, 28.
PRIMAKOFF, H. and HOLSTEIN, T., 1939. *Phys. Rev.* **55**, 1218.
RARITA, W. and SCHWINGER, J., 1941 a. *Phys. Rev.* **59**, 436.
RARITA, W. and SCHWINGER, J., 1941 b. *Phys. Rev.* **59**, 556.
ROSE, M. E. and BETHE, H. A., 1937. *Phys. Rev.* **51**, 205 and 993.
SACHS, R. G., 1939. *Phys. Rev.* **55**, 825.
SACHS, R. G., 1940. *Phys. Rev.* **57**, 194.
SCHMIDT, TH., 1937. *Z. Phys.* **106**, 358.
SCHULER, H., 1937. *Z. Phys.* **107**, 12.
SCHULER, H. and KORSCHING, H., 1937. *Z. Phys.* **105**, 495.
SLATER, J. C., 1929. *Phys. Rev.* **34**. 1293.
SLATER, J. C., 1930. *Phys. Rev.* **35**, 509.
SLATER, J. C., 1931. *Phys. Rev.* **38**, 237.
SPAIN, B., 1940. *Proc. Roy. Soc.* A, **174**, 523.
TELLER, E., 1934. *Eucken-Wolf's Hand- und Jahrbuch der Chem. Physik*, **9** (ii), 43–188.
TYRELL, W., 1939. *Phys. Rev.* **56**, 250.
VOLKOFF, G. M., 1942. *Phys. Rev.* **62**, in press.
WAY, K., 1939. *Phys. Rev.* **55**, 963.
WEFELMEIER, W., 1937. *Z. Phys.* **107**, 332.
WEIZSACKER, C. F. v., 1935. *Z. Phys.* **96**, 431.
WEIZSÄCKER, C. F. v., 1936. *Naturwissenschaften*, **24**, 813.
WHEELER, J. A., 1937. *Phys. Rev.* **52**, 1083.
WICK, G. C., 1934. *Nuovo Cim.* **11**, 227.
WIGNER, E., 1931. *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren* (Braunschweig : Vieweg u. Sohn).
WIGNER, E., 1933. *Phys. Rev.* **43**, 252.
WIGNER, E., 1937 a. *Phys. Rev.* **51**, 106.
WIGNER, E., 1937 b. *Phys. Rev.* **51**, 947.
WIGNER, E., 1939. *Phys. Rev.* **56**, 519.
WIGNER, E., 1940. *Bicentennial Symposium, University of Pennsylvania.*
YUKAWA, H., SAKATA, S. and TAKETANI, M., 1938. *Proc. Phys.-Math. Soc. Japan*, **20**, 319.

PHOTOMETRY

By H. BUCKLEY,
National Physical Laboratory

§1. INTRODUCTION

The past eight years since the last *Report* on this subject was published (Buckley, 1934 a) have seen considerable progress in photometry—that branch of science which is concerned with the measurement of radiant energy in terms of its light-producing effect. This progress has largely been brought about by the necessity of measuring the light-giving properties of luminous sources more and more accurately. The introduction of tungsten vacuum and gas-filled lamps produced problems for the photometric laboratory which were relatively simple. The recent introduction of discharge lamps having very discontinuous spectra, and the importance given to the properties of coloured glasses arising from their ever-increasing use in land, sea and air traffic, have lately confronted the photometric worker with the problem of heterochromatic photometry in its most difficult aspects. The problem has not appeared as an academic one, but as one of great practical importance demanding a practical solution.

There is very little in the way of theoretical background to photometry, which presents itself as the problem of assessing light in relation to human beings whose response to it is very complicated, varying considerably from individual to individual and also with the intensity and conditions of stimulation. Thus photometry is purely a metrological science, and its problems are similar to those of all meteorological sciences.

These problems of photometry are as follows:—

(*a*) The primary standard of luminous intensity (candlepower) by means of which the photometric unit (viz. the candle) is produced and maintained.

(*b*) The definition of the photometric scale, i.e., the definition of the standard observer in terms of whom photometric measurements are expressed when these measurements involve colour differences. The standard observer is defined in terms of the properties of the average eye by means of the visibility curve of the average eye or the luminosity curve, as it is now called.

(*c*) The general problem of heterochromatic photometry when the photometric scale has been defined. This problem is really the forcing of the methods of visual or physical photometry to give results consistent

with the defined properties of the average eye or standard observer when the observers or physical receptors used have properties different from those of the average eye or standard observer.

It is the purpose of this ***Report*** to describe and discuss the progress which has been made in these problems during the past ten or more years.

§2. THE PRIMARY STANDARD OF LUMINOUS INTENSITY

Towards the end of the last century the necessity of measuring the luminous intensities (candlepowers) of gas burners and carbon lamps arose, and consequently luminous flame standards, such as the Pentane lamp, Hefner lamp and Carcel lamp, were developed and their luminous intensities defined under standard conditions.

Experience showed that these standards were not entirely satisfactory, and led to the use of carbon filament lamps, which were found to have sufficiently stable characteristics to enable them to be used as photometric standards. Ultimately, in 1909, comparisons of the unit maintained by means of sets of carbon lamps in England, France and the United States resulted in agreement on the unit adopted in these countries. Germany was not a party to this agreement, but comparisons showed that one Hefner candle was approximately 0·9 of the agreed unit. The International Commission on Illumination, in 1921, agreed to adopt this unit, which became known as the international candle. Germany was not represented at this meeting and continued to use the Hefner candle. The international candle has since been maintained at the national laboratories of England, France and the United States by means of groups of carbon lamps. With a unit maintained in this way it is always possible for it to change with use. It therefore became desirable that the unit should be capable of reproduction even if the lamps maintaining it were destroyed.

Various attempts to produce a fundamental primary standard had been made from time to time. Violle, in 1879, proposed that one square centimetre of melting platinum should be adopted as a primary standard having a luminous intensity of 20 candles. It actually was the nominal legal French standard from 1889 to 1900. Petavel, in England, did a great deal of work on this standard and showed that it was not sufficiently reproducible. Later, Waidner and Burgess (1908) put forward the somewhat modified proposal that the primary standard should be a specified area of a black body maintained at the melting point of platinum. Ives (1924 a, 24 b) found that the brightness of the interior of a platinum cylinder, acting as a black body, was, at the instant of melting, 55·4 international candles per cm^2. Brodhun and Hoffman (1926), using porcelain black bodies, obtained a value of 65·2 Hefner candles or 58·3 international candles per cm^2.

More recently Wensel, Roeser, Barbrow and Caldwell (1931), of the Bureau of Standards, investigated the Waidner-Burgess standard, which was realized by melting platinum in a high-frequency induction furnace. A thoria tube

about 2·5 mm. in diameter was immersed in an ingot of platinum to form a black body. The brightness of the aperture of the tube was determined by focusing it, by means of a lens and right-angled prism, on the screen of a photometer. As the result of a hundred or more melts and freezes on two ingots, it was found that the brightness of the black body at the freezing point of platinum was 58·84 international candles per cm^2. The brightness during freezing was found to be more reproducible than that during melting. This practical realization of the Waidner-Burgess proposal seemed so promising that in 1933 the Advisory Committee on Electricity of the International Committee on Weights and Measures,* which had been asked to consider photometric units and standards, recommended national laboratories to determine the brightness of the black body at the freezing point of platinum in terms of their existing units.

In France, Ribaud (1933 a, 33 b); at the University of Strasburg, obtained the value of 58·78 candles per cm^2, while Buckley and Barnett (1933, 37), at the National Physical Laboratory, obtained the value 59·0 candles per cm^2. Thus it was possible to conclude that the brightness of the black body at the freezing point of platinum was 58·9 candles per cm^2, to an accuracy of at least 0·2 %, and the way was paved for the international adoption of a primary standard of light which would serve to maintain the international candle.

Meanwhile trouble had occurred in connection with the unit maintained at other colour temperatures than that of carbon lamps. The various national laboratories, on the introduction of tungsten vacuum and tungsten gas-filled lamps, each independently prepared standards of these types. They were derived, by photometric comparison, from the carbon-lamp standards, but no definite method of dealing with the colour differences involved had been adopted. It was well known that the results of photometric comparisons involving a colour difference varied from observer to observer and with the conditions of field size and brightness used in the comparison. In stepping up from carbon-lamp colour to tungsten-lamp colour, each laboratory had carried out the comparisons in its own way, and the agreement between their tungsten-lamp standards was by no means satisfactory.

At a meeting of the International Commission on Illumination in 1928†, it was stated that, at that time, the ratio of the International to the Hefner candle, instead of being constant, varied with colour temperature, as follows:

Colour temp.	Ratio Int. candle/Hefner candle
2000° K.	1·11
2360° K.	1·145
2600° K.	1·17

In making the step from 2000° K. to 2600° K. and higher, none of the laboratories could state that the method used was such as would ensure that the values of standards at the various colour temperatures were consistent with the properties

* *Com. int. Poids. et Mes., Proc. verb.* **15**, 200, 1933.
† *Com. int. d'Eclair.* vii^e Sess. p. 1168, 1928.

of the average eye, which had been defined in 1924 by the International Commission on Illumination* in terms of an internationally agreed luminosity curve.

The International Committee on Weights and Measures officially adopted the I.C.I. visibility curve at its meeting in 1933†, and at its meeting in 1937‡ recommended that all derived photometric quantities should be consistent with this luminosity curve. It became apparent that if the new platinum standard were adopted, with its brightness defined as 58·9 candles per cm^2, adjustments would have to be made in the values assigned to existing tungsten lamp standards operating at 2360° K. and higher temperatures up to 2800° K. The proposal was then made to the International Committee by the National Laboratories of the United States, France and Great Britain§ jointly, that instead of adopting the platinum standard with a brightness of 58·9 international candles per cm^2, the unit itself should be changed so that the brightness of platinum in terms of the new unit would be 60 *new* candles per cm^2. This would increase the numbers expressing the candlepower of the existing carbon standards, and similarly, those expressing the candlepowers of standards operating at 2360° K. and 2800° K. As it was known that the standards at 2360° K. were not consistent with the visibility curve, and should have their candlepowers reduced by about 1·5%, the adoption of this proposal would mean that the candlepowers of these standards in international units should first be reduced by about 1·5% and then increased by 1·8% to express them in terms of the new unit. This would leave the numbers expressing their candlepowers in terms of the new unit practically the same as the existing numbers (which erroneously expressed them in terms of the international candle). Some adjustment would still be necessary to the numbers expressing the candlepowers of gasfilled standards in terms of the new candle, but in view of the method used in deriving them, some adjustment would have to be made in any case. The adoption of this proposal would make the actual changes in the numbers expressing their candlepowers less than otherwise would be the case.

It was therefore recommended at a meeting of the Advisory Committee on Photometry of the International Committee of Weights and Measures in 1937‖, that the new unit should be adopted from January 1st, 1940, and that it should be called the "new candle". Germany was a party to this decision, which meant that the Hefner candle would no longer be used. It was further decided that before that date it would be desirable to have an international inter-comparison of lamps maintaining the unit at the colour temperature of freezing platinum (2046° K.) and at the colour temperature 2360° K.

Accordingly lamps were sent by five national laboratories and compared with each other at the National Physical Laboratory. The results of the com-

* *Com. int. d'Eclair.* vi^e^ Sess. 67, 1924.
† *Com. int. Poids et Mes., Proc. verb.* **15**, 203, 1933.
‡ *Com. int. Poids et Mes., Proc. verb.* **16**, 237, 1937.
§ *Com. int. Poids et Mes., Proc. verb.* **18**, 243, 1937.
‖ *Com. int. Poids et Mes., Proc. verb.* **18**, 236, 1937.

parison are given in the following table, which shows the relative values of the new unit as realized at the different laboratories—the mean of the units of the different laboratories at each colour temperature being taken as unity.

Laborstory	Value of new unit	
	at 2046° K.	at 2360° K.
National Bureau of Standards	0·9990	0·9994
Phys. Tech. Reichsanstalt	0·9964	0·9931
Lab. Cent. d. Electricité	1·0061	0·9972
Electrotech. Lab., Tokio	0·9968	1·0084
National Physical Laboratory	1·0016	1·0018

These results were considered by the Advisory Committee on Photometry at its meeting in 1939,* and the recommendation that the new unit should be introduced in 1940 was confirmed.

The standards operating at colour temperature 2360° K. were derived and compared with those operating at 2046° K. by combining the latter with a blue glass which made the colour of the resultant radiation very closely the same as that from the higher temperature standards. The energy distributions were, however, somewhat different. The integral transmissions of the blue glasses were computed from spectrophotometric measurements of their transmission factors throughout the visible spectrum, the spectral energy distribution of a black body at 2046° K. and the I.C.I. luminosity curve. Since the method eliminates colour differences and the calculated integral transmission factors of the blue glasses are necessarily consistent with the I.C.I. luminosity curve, the standards of colour temperature 2360° K. derived should also be consistent with the I.C.I. luminosity curve. It is desirable, however, that the standards of colour temperature 2360° K. obtained in this way should be shown to be consistent with the luminosity curve by an independent method also involving the luminosity curve. This has been done by Wensel, Roeser, Barbrow and, Caldwell (1934) of the Bureau of Standards, who have shown that their standards operating at colour temperatures 2046° K. (freezing point of platinum) and 2727° K. (freezing point of iridium) are closely consistent. The method, in principle, consists in determining the luminous intensities of tungsten lamps which colour-match a black body at 2727° K. in terms of the brightness of this black body. The brightness of the black body at 2727° K. is then compared with that of a black body at 2046° K. when both are viewed through a red glass of known effective wavelength. If the brightness of the black body at 2046° K. is known, then, by means of Planck's radiation law, together with the I.C.I. luminosity curve and the measured ratio of brightnesses at the effective wavelength, the brightness of the black body at 2727° K. can be deduced and hence the luminous intensities of the tungsten standards which colour-match it.

Unfortunately the war broke out before the recommendation of the Advisory

* *Com. int. Poids et Mes., Proc. verb., Com. de Phot.* p. 49, 1939.

Committee could be ratified by the International Committee on Weights and Measures, which alone has the necessary executive powers. Accordingly, no action has been taken by any country to put the new unit into use. Thus this very desirable change in the status of the fundamental photometric standard and unit has had to be postponed until happier times.

§3. THE PHOTOMETRIC SCALE—THE STANDARD LUMINOSITY CURVE

The second problem is that concerned with the measurement of sources differing in colour from the primary standard. This is the problem of heterochromatic photometry. It arises from the fact that photometric measurements on light sources depend on the observer who makes them. Even for one observer they depend on the conditions under which they are carried out. To solve this problem it is necessary to specify the visual characteristics of the observer in terms of whom photometric measurements are to be expressed. This has long been recognized, and was solved, in principle, in 1924, when the International Commission on Illumination * adopted a standard visibility or luminosity curve to represent the visual characteristics of the average eye or standard observer for photometry. This curve represents the relation between wavelength and luminosity per unit of radiant energy at each wavelength. It was based largely on the work of Gibson and Tyndall (1923), who used the step-by-step method, in which neighbouring portions of the spectrum are compared and large colour differences avoided. Field brightnesses were fairly high and a small 3° visual field was used. The authors concluded, from a comparison of their results with those of other observers, that there was no certain difference between the luminosity curve determined by the step-by-step method and by the flicker method, if both were carried out with a high field brightness and a small visual field, i.e. if the photometric conditions recommended by Ives (1915 a, 19) were observed.

As stated previously, the International Committee on Weights and Measures approved this luminosity curve at its meetings in 1933 and 1937, when it decided that photometric quantities, if they involved measurements in the presence of colour differences, should be determined by procedures based on the luminosity curve. This resolution does not, however, prescribe the conditions under which heterochromatic photometry is to be carried out. It only determines the photometric scale which shall be used, i.e. that of the standard observer possessing the agreed luminosity curve.

Considerable confidence has been gained from experience in the use of the agreed luminosity curve, which came to be extensively used in photometry In particular, the filter method of carrying out heterochromatic photometry became widely used. In this method, instead of making comparisons in the presence of a colour difference, a colour filter is used in conjunction with a source

* *Com. int. d'Eclair.* vi^e^ Sess. p. 67, 1924.

of known colour temperature. The transmission factor of this filter for the radiation from the source is calculated from the spectral transmission of the filter, the energy distribution of the source and the luminosity curve. Thus the transmission of a coloured filter for light of colour temperature T is given by

$$T = \int_0^\infty E(\lambda T)K(\lambda)\tau(\lambda)d\lambda \Big/ \int_0^\infty E(\lambda T)K(\lambda)d\lambda,$$

where $E(\lambda T)$ is the rate of energy emission of a black body at temperature T per unit wavelength at wavelength λ, $K(\lambda)$ is the luminosity factor and $\tau(\lambda)$ the transmission factor of the filter at wavelength λ.

In some cases the production of a colour match by this method produces an energy match between the sources compared, so that the photometry is reduced to homochromatic photometry, and so is independent of the observer carrying out the comparison. In general, however, even perfect colour matches will not at the same time produce energy matches, so that there is some dependence on the visual properties of the observer, and, in consequence, some uncertainty in the final result taken as the mean of the results obtained by a group of observers.

That the I.C.I. luminosity factors were satisfactory appeared from the wor. of Crittenden and Taylor (1929) and that of Buckley (1934 b). In their experiments the integral transmission factors of a number of coloured glasses were determined by spectrophotometry and calculations based on the luminosity curve and also experimentally by means of the flicker photometer used under the conditions recommended by Ives (1915 a, 19). The experimentally obtained results were corrected, as proposed by Ives (1915 a, 19), to those of the standard observer (whose Y/B ratio* is unity) from plots of the observer values against the Y/B ratio of the individual observers.

In recent years the validity of the luminosity curve has been questioned by a number of German investigators. These investigators attached considerable importance to the use of small visual fields of the order of 1°, with the object of making photometry depend on pure cone vision—the Purkinje effect being believed to be non-existent for small visual fields surrounding the fovea.

Arndt (1936), using the green and yellow lines in the mercury spectrum, compared their energies by means of a calibrated rectifier cell and their luminosities by means of a flicker photometer (1·5° field, 30 metre candles illumination). The observed ratio of the visibility factors was found to be 1·035, instead of 0·903 derived from the luminosity curve. Dresler (1937) confirmed this discrepancy by measurements which included the sodium D line and the green line of the mercury spectrum. He found that the ratio of their luminosity factors was 0·94

* The Y/B ratio of an observer is the ratio of the transmissions of two solutions, one yellow and one blue, obtained under certain standard conditions by means of a flicker photometer. The average value of the Y/B ratios of a large group of observers, in these circumstances, is unity, the individual values lying mainly between 0·90 and 1·12. It has been found that photometric results, involving colour differences such as are produced by coloured glasses and by variation in colour temperature of lamps, plot linearly with respect to the Y/B ratios of the individual observers.

for a 1·2° field and 0·99 for a 7° field, whereas the luminosity curve gave 0·777. He also determined the transmissions of four coloured filters by means of the flicker photometer and also by means of spectrophotometry and calculation. He then showed that the luminosity curve most consistent with his observed results on the filters and spectrum lines was similar to the I.C.I. luminosity curve, but displaced towards the red end of the spectrum with its maximum at 5650–5700 A. as compared with 5550 A. As a result of these measurements the Germans in 1937 suggested that the I.C.I. curve was in error and proposed that a re-determination of it, which should be characteristic of cone vision, should be undertaken.

Since 1937 a number of re-determinations and checks of the luminosity curve have been made. A very complete re-check by Jainski (1938) using a step-by-step flicker method (field size 1·2° and field illumination 20-80 m.c.) also confirmed the shift in the luminosity curve. For a 6° field he found his results approximated to the I.C.I. curve. Jaggi (1939) has determined, by the flicker method (field size 1·2°, field illumination 50 m.c.), the integral transmission of a series of coloured filters and, using a method developed by Konig (1938), has deduced the luminosity curve most consistent with the results. This curve also showed a shift towards the red as compared with the I.C.I. curve, but not as much as that suggested by Dresler's results.

With results at variance with those of the German and Swiss observers, Preston (1938), and also Forsythe, Barnes and Shrider (1938), have made measurements of the ratio of the luminosity factors of the yellow and green lines of the mercury spectrum. Preston obtained 0·903 for a group of 16 observers, the average of whose Y/B ratios was 1·003, while Forsythe and his co-workers obtained 0·923, compared with the I.C.I. value of 0·903. Both these determinations were carried out with a visual field of 2° and field brightness of about 25 m.c. Preston also showed that no significant change occurred if a 1° field were used. Later, Preston (1939) repeated his work and extended it to include a comparison of the luminosity factors of the main lines in the spectra of mercury, cadmium and sodium lamps and found values which confirmed the I.C.I. curve, except in the region 5600–6300 A., where his results were higher by amounts up to about 10%.

The German observers seem to have ignored the value of determining the Y/B ratios of the observers used in their work. If their new curve were correct, it would mean that the mean Y/B ratio of a sufficiently large group of observers should be about 1·10, instead of 1·00 on the basis of the I.C.I. curve. Actually, the only information available shows that, with 2° fields and 25 m.c. illumination, large groups of observers at the National Physical Laboratory and the National Bureau of Standards obtained mean values of the Y/B ratio not far from unity.

The question of altering the I.C.I. luminosity curve was considered at the 1939 meeting of the International Commission on Illumination, when the conclusions put forward by Urbanek and Ferencz (1939), in a very complete

report on visual photometry, were agreed to. These conclusions were: (i) There exists no average luminosity curve which is a function solely of wavelength; (ii) it is impossible to eliminate completely the effect of variables other than wavelength (the use of pure cone vision is no remedy); (iii) the average luminosity curve must be defined arbitrarily to a certain extent. In consequence the Commission decided that there was no need to alter the I.C.I. luminosity curve, and that all photometric measurements should be consistent with it.

By " consistent with the luminosity curve " is meant that the result obtained by the photometric procedure adopted should be consistent with the evaluation of the ratio

$$\int_0^\infty E'(\lambda T)K(\lambda)d\lambda \Big/ \int_0^\infty E_0(\lambda T)K(\lambda)d\lambda,$$

where the denominator refers to the primary standard and the numerator to the test source or its equivalent. Incidentally, the experimental determination of the ratio of integrals of this type, when $E'(\lambda T)$ is often unknown by means of visual or physical apparatus whose properties are not identical with $K(\lambda)$, constitutes the problem of heterochromatic photometry.

The discussion and work carried out in consequence of the German criticisms have not, however, been without value. Considerably improved technique was used by the various investigators. It has also emphasized the fact that photometric results are somewhat arbitrary in character. They do not now claim to be the results which would be obtained by an observer who possesses the average eye, since the average eye varies with the circumstances under which it is deduced. They are the results which would be obtained by a hypothetical standard observer defined by the I.C.I. standard luminosity curve. So long as the results thus obtained do not seriously differ from the results of experience, we are content to take his results as the useful ones for photometric purposes.

During the progress of these experiments the possibility of dietetic influence as a possible cause of variations in an individual's luminosity curve was put forward by Preston (1939), since it had become known that deficiency of vitamin A has a surprisingly large effect in raising the absolute threshold of vision of the dark adapted eye. Dresler (1940) has since recorded a seasonal change in the ratio of the luminosity factors of the 5893 A. and 5461 A. lines, which is possibly connected with seasonal variation in the vitamin A content of the body. The ratio was found to be about 6% lower in summer than in winter.

§ 4. HETEROCHROMATIC PHOTOMETRY

(a) *Visual photometry*

In the early years of the nineteen-thirties, the modern gaseous discharge lamps made their appearance and presented the photometric worker with the problem of heterochromatic photometry in its most acute form, since the radiation from some of these is confined to a few wavelengths only. The principal gases used in these lamps are neon, sodium vapour and mercury vapour

(high or low pressure). Earlier papers on the photometry of these sources confined themselves to describing colour filters which could be used with tungsten lamps to obtain colour matches. Among these are those by Paterson (1932), Dziobek and Reeb (1933), Yamauti (1931) and Fabry, Roux and Perrin (1929), Dresler (1938) also gives particulars of German glasses suitable for use in the photometry of neon, sodium and mercury lamps.

Winch and Palmer (1934) have described a method of selecting suitable colour filters which is equivalent to determining that thickness of a suitable glass which gives results as independent as possible of the Y/B ratio of the observer. More recently, Little and Estey (1935) have described a more fundamental method inasmuch as they determine the effect of thickness and lamp-colour temperature on the position of the point on the colour triangle which represents the colour of the transmitted light. The thickness of filter and colour temperature most suitable are those for which the colour point of the combination is the nearest to that of the source. They give particulars of filters or filter combinations which colour match sodium and mercury lamps.

Other developments in connection with discharge lamps have been concerned with the production of sources whose light is distributed more or less continuously throughout the spectrum. This is accomplished by the use of fluorescent materials on the inside of the bulbs of mercury-vapour lamps. In this way the invisible ultra-violet radiation of the mercury discharge is used to add fluorescent light to the visible radiation from the discharge. The spectrum tends to approach a continuous one, so that the photometric difficulties are not so severe.

Early results showed that different laboratories differed considerably in the values they obtained from lumen measurements on high-pressure mercury lamps (Forsythe, 1936). It is evident that the use of colour-matching filters only removes the psychological difficulty of deciding what is a brightness match. The inherent differences between observers who are classed as normal still remain. Buckley (1934 b), and also Winch and his associates (1934), have described measurements on gaseous discharge lamps carried out by different methods, including the flicker method under the Ives' conditions and also direct comparison by the full-colour difference method and by the colour-filter method, each with 2° and 10° visual fields. Buckley's results showed that all methods with a 2° field gave substantially the same results. When large field sizes were used, field-size effects were apparent with the full-colour difference method and also with the mercury lamps when the colour-match method was used. In this case it was found that the mean of the result of a group of twelve observers was about 8% higher with a 2° field than with a 10° field. In consequence he recommended that the 2° field should be used because, as stated previously, the luminosity curve used to calculate the integral transmissions of the filters was considered to apply more particularly to a small portion of the retina surrounding the macula than to a large area of the retina. The use of the 2° visual field has since become general in the measurement of discharge lamps both in England and America, and, it is believed, elsewhere.

Forsythe, Barnes and Shrider (1938) have confirmed the existence of definite difference in results with the colour-match method of photometry of mercury lamps, depending on the field size. Their work and that of Buckley (1934) showed that even with the colour-match method, the individual observers in a group of about ten can be expected to have a spread in their results of the order of 10% or more in the photometry of high-pressure mercury lamps. This spread in the results does not correlate well with the Y/B ratios of the observers concerned, so that Ives' method of plotting results against Y/B ratios of observers and deducing the result corresponding to a Y/B ratio of unity is not satisfactory. A more refined test or weighting of observers is necessary. Such a test is not at present available. All that can be done is to relate the results of individual observers to the mean of a large group of the order of 25 or more, if the highest accuracy possible is desired. It does not therefore seem possible to be certain of candlepower or lumen values of these lamps to an accuracy much better than about 2%. According to König (1939), the situation is somewhat the same for the sodium lamp. The yellow colour-matching filters usually employed transmit all but the blue end of the spectrum, so the energy difference in the spectra of radiations compared is very great. For these lamps, filters similar to that of Gibson (1935), which transmits a very narrow region of the spectrum near wavelength 5600 A., are desirable. None centreing about 5893 A. has as yet been described.

The effect of selectivity of the paint used in photometric integrators has been investigated by Buckley (1935) and by Benford (1935). The former showed that, with suitable methods of using the colour-match filters, measurements of the luminous outputs of neon, sodium and mercury lamps should not be in error by more than 1% due to reasonable amounts of permitted selectivity. The latter described an experimental method of determining *in situ* the " apparent reflection factor " of the sphere paint for various light sources and consequently the amount of error introduced by the selectivity.

(b) *Physical photometry*

(i) *Photo-emissive cells.* Considerable attention has been paid during the past decade, and particularly in its earlier years, to the application of photoelectric cells of the emission type to photometry. Bulk metal cells have been largely replaced by thin film cells in which a mono-molecular layer of metal is deposited on a composite surface usually consisting of an oxide layer on a metallic film. The spectral response curves of the potassium, rubidium and caesium types of these cells cover the whole of the visible spectrum, so that, theoretically at any rate, their spectral response curves can, by means of appropriate filters, be modified to agree with the I.C.I. luminosity curve. The work function of caesium is, however, low, so that for caesium cells the thermionic current is rather large and is liable to vary considerably with temperature; they are, therefore, not as stable as the other types. Vacuum cells appear to be almost

exclusively used for photometry on account of their inherently greater stability than the more sensitive gas-filled cells in which ionization by collision occurs. For approximately homochromatic photometry, the fact that the spectral sensitivity is different from that of the I.C.I. luminosity curve is not of great importance. In heterochromatic photometry, however, it becomes of supreme importance.

Most, if not all, lamp factories nowadays control their production of tungsten lamps by means of lumen output measurements, using photometric integrators and corrected photocells. The correction is obtained by means of filters which are empirically selected or adjusted, so that the combination of cell and filter gives the same results as those obtained by the standard observer. Preston and McDermott (1934 a) have described methods of correcting cells by means of liquid filters. Barbrow and Meyer (1932) have determined the photometric characteristics of tungsten lamps by means of a gas-filled caesium cell suitably corrected and frequently checked by measurements made on sub-standards operating at colour temperatures from 2400° to 2800° K. whose values were determined visually.

One of the desirable properties of a physical measuring instrument is that its response should be proportional to the stimulus. Some controversy at one time ranged over the characteristics of photocells in this respect. There is a certain amount of fatigue noticeable in the use of these cells, the photo-current altering slightly immediately after exposure. There appears to be a connection between fatigue and non-linearity of response. Cells do, however, exist which show practically no fatigue and are linear to a high degree of accuracy over a considerable range of illumination. It is not safe, however, to assume that any given cell is satisfactory without verification. In some cases, as shown by Preston and McDermott (1934 b), the departure from linearity may vary with wavelength, so that the cells exhibit a kind of Purkinje effect. Boutry and Gillod (1939) have recently described a caesium cell in the construction of which particular attention was paid to the elimination of all causes which might lead to non-linearity of response. It was specially designed for photometric measurements of the highest precision.

Winch (1930) and Winch and Harper (1932) have described photoelectric equipment and methods which are used in lamp factories. Methods in use at the N.P.L. have been described by Harrison (1929) and by Walsh (1933). Valve amplification is usually employed, and the circuit which appears to be most favoured is that first used in photometry by Sharp and Smith (1928). In this circuit the photo-current passes through a high resistance, the voltage drop down which is communicated to the grid of a thermionic valve. This valve forms one arm of a Wheatstone bridge, the other arms of which consist of two resistances and a second thermionic valve closely matching the first in electrical properties (variation of anode current with filament volts and grid volts). The bridge is balanced with the photocell in the dark and then when illuminated.

The balance is obtained by neutralizing the voltage drop down the high resistance by means of a potential derived from a potentiometer. The magnitude of this neutralizing potential is proportional to the illumination of the cell.

The general utility of such an arrangement depends very much on the properties of the high resistance. Since the photo-electric cell is a device which gives current independently of the resistance in series with it, the higher the resistance the more sensitive the circuit becomes. Resistors of the order 10^{10} ohms, which are very stable and have small temperature coefficients, can now be obtained commercially. With very high resistances insulation troubles begin to appear and electrical screening become important. Winch (1940) has avoided these difficulties by devising a most compact arrangement of cell, electrometer valve and high resistance, all of which are enclosed *in vacuo*. Preston (1936 a) has described a photoelectric photometer circuit using resistances up to 10^9 ohms and capable of measuring photo-currents as low as 10^{-14} amp. to an accuracy of about 1% without such drastic treatment.

Secondary emission cells have recently been developed. Here amplification of the primary photo-current is accomplished by using the primary electron emission to produce secondary electrons by bombardment of suitable surfaces. In favourable circumstances, each primary electron may produce ten secondary electrons. The possibility exists of constructing these secondary cells so that several stages of amplification result. So far, secondary emission cells do not appear to have been successfully used for photometric purposes.

(ii) *Rectifier cells*. The most important single event affecting photometry during the past ten years has been the discovery of the barrier layer or rectifier photocell. This cell consists of a layer of a semi-conductor between two metallic plates. The upper plate may be either a grid or a transparent sputtered or evaporated film of gold or platinum. In these cells the barrier layer between the semi-conductor and the metallic film behaves as a rectifier.

They also have the property of converting incident visible radiant energy into the electrical energy of an electron current, without the necessity of using a battery, when the upper and lower plates are electrically connected. Preston (1936 b) has given an account of their manufacture and use in photometry. The semi-conductors used are usually selenium or copper oxide. The spectral response curve of the latter is not very different from the I.C.I. luminosity curve. That of the former has its maximum at about $0{\cdot}56\,\mu$, but is considerably wider than the I.C.I. curve and continues into the ultra-violet and infra-red portions of the spectrum. Barnard (1935) has described a method of making rectifier cells using a mixture of selenium and sulphur as the semi-conductor.

It was soon discovered that the photo-current developed by these cells was not always proportional to the illumination. In fact, the higher the resistance of the measuring system included between the plates the more the current deviated from proportionality with the illumination. These early experiments also showed that the cell behaved as if it included a shunt resistance which varied

with the illumination: in other words, some electrons liberated at the upper surface of the semi-conductor get back inside the cell. If the resistance of the measuring instrument were zero, then all the current would go through the instrument and, in consequence, the photo-current would be proportional to the illumination. For instruments of very low resistance, and for the method of measurement described by Campbell and Freeth (1934), in which the cell is effectively short-circuited, whatever the resistance of the measuring instrument, the cells do behave approximately linearly. Experiments by Buchmüller and König (1937) showed that this is not always the case, and that the departure from linearity could be either positive or negative and was not necessarily zero under short-circuit conditions. Atkinson, Campbell, Palmer and Winch (1938) have investigated the effect very exhaustively and showed that the effects could be explained by assuming the presence of a series resistance as well as a shunt resistance inside the cell, and that these two resistances did not vary in the same way with illumination.

The first use of these cells was for portable illumination photometers in conjunction with microammeters. The cells suffer one objectionable disability, inasmuch as they do not obey the cosine law. With some of the earlier cells the current for an angle of incidence of 70° might be of the order of half that for the same illumination under normal incidence. Methods of correcting cells for this defect have been described by Barnard (1936). Barnard's method consisted of placing an aperture a short distance in front of the cell. The effect for oblique incidence is to shift the effective area of cell towards the periphery, where, for some cells, the sensitivity is higher than at the middle. Considerable improvement has been made by using matt lacquers on the top surface of cells in these last few years, so that the deviation from the cosine law is in many cases not more than about 20% at angles of incidence as high as 80° (*Brit. Std. Spec.* 1936). Portable illumination photometers using rectifier cells are now available commercially for measurements covering the range 0–500 f.c., which is divided into a number of ranges, either by electrical shunts or diaphragms. The lowest range on one model covers 0–0·3 f.c. with a scale about 4" long, so that measurements of street lighting can be made with reasonable accuracy.

Rectifier cells are not as stable as emission cells and exhibit more fatigue, particularly for the red portion of the spectrum. Recent examples show much improvement over earlier ones in all respects.

As with emission cells, it is possible by means of colour filters to correct the sensitivity curve of rectifier cells to approximate agreement with the I.C.I. visibility curve. König (1934, 37) has done considerable work on this problem and has constructed several " precision artificial eyes ". In particular he has shown how it is not necessary to have an actual filter to effect the correction. The correct results can be obtained by means of a linear combination of the results obtained by the use of the individual members of a suitable series of filters, each being suitably weighted. Barbrow (1940) has shown how rectifier

cells, whose correction to the I.C.I. luminosity curve is not perfect, may be used for precision photometry. His method consists in the use of a substitution method in which the unknown source is compared with a known source which, by combination with filters of known transmission, colour matches it.

(iii) *Dispersion-template method.* The last few years have seen considerable revival of interest in a method of physical photometry investigated by Ives in 1915. In this method, light from the source under test is dispersed by a prism, and in the plane of the spectrum a template or mask of suitable shape is placed. The spectrum is then recombined on a thermopile. The template is cut so that its shape is the same as the luminosity curve of the eye as modified by the dispersion of the prism. Ives (1915 b) showed that this method was practicable, but that it required a very high order of galvanometer sensitivity and could only be used with very high illuminations of the slit of the spectroscope.

König (1934) has described a simplification in which the modification of the response of a selenium rectifier cell, which he used instead of a thermopile, took place in two steps. The first step was a rough adjustment by suitable colour filters, with the final relatively small adjustments being made in the template. The higher sensitivity of the rectifier cell as compared with the thermopile promises to make this method more practical than it was with the thermopile as receiver. More recently still the method has been used with vacuum emission cells of the potassium and rubidium types by Voogd and Went (1939) and by Winch (1940). In each case the template is empirically adjusted *in situ* at a number of wavelengths, so that the galvanometer deflections obtained at the various wavelengths are proportional to those calculated from the colour temperature of the calibrating source used. In this way the selective absorption of the instrument or of the paint and window of the integrating sphere is compensated for in the adjustment of the template. Voogd and Went present evidence that the integral transmissions of coloured filters determined with their instrument agree very closely with those determined by spectrophotometry and calculation. Winch shows that determinations of the luminous outputs of modern discharge lamps agree with the visually determined values to about the accuracy which can be claimed for the visual values.

(iv) *Thermopiles and bolometers.* Just as there has been a revival of interest in the dispersion-template method, so there has been a revival of interest in the physical photometer solution method of Ives and Kingsbury (1915), in which a solution whose spectral transmission curve is proportional at every wavelength to the I.C.I. visibility curve is used in front of a thermopile.

The original liquid filter devised by Ives and Kingsbury suffered from the disadvantage that it transmitted practically no light below wavelength 0·46 μ. Gibson, Teele and Keegan (1938, 39) have described an improved solution which is used in combination with a plate of yellow uranium glass. Teele (1941 a, 41 b) has described equipment with which the improved luminosity filter is used to produce a precision physical photometer. Small electromotive forces of the order of 10 microvolts have to be measured to an accuracy better

than 1%; so that a special potentiometer had to be designed. This has been described by Teele and Schuhmann (1939). It is claimed that the luminous transmission of highly selective signal glasses can be obtained with considerable accuracy.

König (1937) has prepared a set of twelve glass filters which, when used with a thermopile, produces a practicaily perfect "physical eye". He also showed how the filters could be used without undue reduction of the light transmitted This he did by assigning weights to each filter, or combination of small number of filters, so that the final result was obtained by a linear combination of the results obtained with the separate filters or filter combinations. This method enabled the integral transmissions of coloured filters to be obtained to an accuracy of about 1%. He points out that the method is only suitable for well-developed laboratories and not for routine work.

Moon and Mills (1934) and Moon (1935) have described a bolometer whose receiving elements carry a.c. instead of the customary d.c. In these circumstances irradiation of one of the elements produces a periodic e.m.f. across the bridge which can be amplified in the usual ways. Combined with an Ives', or similar, solution the bolometer becomes a physical photometer.

More recently Moon and Steinhardt (1938) have described a dielectric bolometer in which the receiving elements are insulating materials. The large change in impedance with temperature of these materials enables considerably higher sensitivity to be obtained than with the metallic bolometer. So far no application to photometry seems to have been made with either of these instruments.

§ 5. PHOTOMETRY AT LOW LEVELS OF BRIGHTNESS

The restricted lighting adopted during war time has made necessary photometry at low levels of brightness. As discussed previously, the object of photometric measurements is to obtain the results which would be obtained by the standard observer defined by the I.C.I. luminosity curve. This it is possible to do (apart from experimental difficulties) by means of corrected physical photometers and also visually by means of the filter method. Normally we are concerned with illuminations and brightnesses which produce comfortable vision. The compromise I.C.I. luminosity curve roughly defines the average observer for these conditions, and, in spite of individual differences, the results obtained are not palpably contradicted by experience in these conditions. Brightnesses which appear equal to the standard observer do not appear intolerably different to real observers.

At low levels of brightness the situation is vastly different, due to the fact that for real observers the luminosity curve depends on the brightness level at which it is deduced. As brightness falls below about one equivalent foot-candle the luminosity curve continuously shifts towards the blue end of the spectrum. At each level of brightness, a real observer has different visual properties. Thus differently coloured surfaces whose brightnesses are adjudged equal on the basis of the standard luminosity curve, i.e. as seen by the hypothetical standard

observer, are often palpably very different in brightness when viewed by real observers. This shift in the luminosity curves of real observers is, of course, responsible for the Purkinje effect, in consequence of which differently coloured surfaces, which to a real observer appear equally bright, do not appear equally bright if the stimulus from each is reduced in the same proportion. That is to say, in the low brightness region, there is for real observers no photometric scale in which the whole is equal to the sum of the parts, i.e. measurement as ordinarily understood is impossible.

Information, however, is often wanted on the " apparent brightness " of materials to real observers as distinct from their brightness as seen by the hypothetical standard observer. All that can be done in these circumstances is to identify brightnesses and to compare them with the brightness of a standard surface whose brightness is fixed by definition. At low levels of brightness the colour sensation tends to disappear, so that the difficulty of obtaining brightness matches is not as great as might be expected. This method has been adopted in Great Britain in connection with the measurement of the apparent brightness of fluorescent materials of low brightness (*Brit. Std. Spec.* 1941). It also has other applications. The standard surface is a white surface of known reflection factor illuminated by a lamp of known candlepower and of colour temperature 2360° K. The brightness of this surface is deduced by the application of the inverse square law, or by the use of a sector disc or other non-selective method of varying illumination. The apparent brightness of materials tested in this way are expressed in terms of what are called " equivalent foot-candles ".

The same difficulty has been met in Germany (Anon., 1940) in connection with A.R.P. lighting, and has been solved in a similar manner. The Germans have introduced two new units for the low brightness region. These are the *skot* and the *nox*. for brightness and illumination respectively. Their magnitudes are such that for sources of colour temperature 2360° K. 1 *skot* $= 10^{-3}$ equivalent metre candles and 1 *nox* $= 10^{-3}$ metres candles.

The problem of photometry at low brightness levels has been discussed by Walsh (1940), while Guild (1932) has dealt with the philosophical aspects of the fact that measurement, as ordinarily understood, is meaningless and impossible in the low brightness region.

§6. MISCELLANEOUS

Stiles and Crawford (1933) have shown that the brightness of an extended source is considerably higher when viewed through a small artificial pupil over the centre of the pupil than when the same artificial pupil is displaced towards, but still remains within, the periphery of the pupil. This phenomenon is referred to as the Stiles-Crawford effect. They have further shown that the magnitude of the effect varies with the wavelength of the light used. With pure substitution methods of photometry, in which the size of pupil used is kept constant, the effect should not cause any difficulty. Collier and Taylor (1938) have pointed

out that considerable errors are possible in some methods of telephotometry, if the size of the exit pupil is not the same when the telephotometer is used with point sources as when it is calibrated with non-point sources. Dziobek (1934) has also confirmed the existence of the effect and called attention to its significance in photometry in certain cases.

A determination of the luminosity curve of radiant energy at low intensities has been made by Weaver (1937). The intensities covered the range of brightnesses from near the dark adapted threshold to about two equivalent foot-candles. At the lowest intensity the wavelength of maximum luminosity was found to be displaced towards the blue, as shown by König (1892) and by Hecht and Williams (1923).

The last few years have seen considerable advances in mathematical photometry, which had remained practically unchanged since the time of Lambert (1760). Beer (1854) considered a number of surface sources, and Lommel (1880, 89) the light emission from volume distributions of emitting (absorbing or non-absorbing) particles. The problem of the inter-reflection of light in various enclosures has recently been investigated by Yamauti (1926, 37) and by Buckley (1927, 28, 34 c). The latter showed that all such problems could be expressed in terms of an integral equation which he solved explicitly in some cases and approximately in others. Whitmore (1939) and Moon (1940) have also dealt with the subject, the latter bringing it into relation with the Schmidt-Hilbert theory of integral equations. Gershun (1932) has studied the emission of a number of volume distributions of light sources such as are encountered in luminous discharge tubes. Although the light field is an electromagnetic field, the modern methods of vector analysis have, until recently, been neglected. Attention was called by Hyde (1907) to the desirability of using these methods, which have now been developed into the theory of the light field by Yamauti (1932) and by Gershun (1936) (see also Moon and Timoshenko, 1939). It has, however, not yet been very successful in solving problems other than some which have already been solved by other methods.

REFERENCES

Arndt, 1936. *Licht*, **6**, 75.

Atkinson, J. R., Campbell, Palmer and Winch, 1938. *Proc. Phys. Soc.* **50**, 934.

Barbrow, 1940. *Bur. Stand. J. Res., Wash.*, **25**, 703.

Barbrow and Meyer, 1932. *Bur. Stand. J. Res., Wash.*, **9**, 721.

Barnard, 1935. *Proc. Phys. Soc.* **47**, 477.

Barnard, 1936. *Proc. Phys. Soc.* **48**, 153.

Beer, 1854. *Grundris des Photometrischen Calcüles* (Berlin).

Benford, 1935. *J. Opt. Soc. Amer.* **25**, 332.

Boutry and Gillod, 1939. *Phil. Mag.* **28**, 163.

British Standard Specifications, 1936. B.S. 667. *Photo-electric Type Portable Photometers.*

British Standard Specifications, 1941. B.S./A.R.P. 18. *Fluorescent and Phosphorescent Paint for A.R.P. Purposes.*

BRODHUN and HOFFMAN, 1926. *Z. Phys.* **37**, 137.
BUCHMÜLLER and KÖNIG, 1937. *Bull. Schweiz. elect. Ver.* No. 5, p. 89.
BUCKLEY, 1927. *Phil. Mag.* **4**, 753.
BUCKLEY, 1928. *Phil. Mag.* **6**, 447.
BUCKLEY, 1934 a. *Rep. Prog. Phys.* **1**, 150.
BUCKLEY, 1934 b. *Illum. Engr., Lond.*, **27**, 118 and 148.
BUCKLEY, 1934 c. *Phil. Mag.* **17**, 576.
BUCKLEY, 1935. *Phil. Mag.* **20**, 745.
BUCKLEY and BARNETT, 1933. *Com. int. Poids et Mes., Proc. verb.* **16**, 856.
BUCKLEY and BARNETT, 1937. *Com. int. Poids et Mes., Proc. verb.* **18**, 247.
CAMPBELL and FREETH, 1934. *J. Sci. Instrum.* **11**, 125.
COLLIER and TAYLOR (W. G. A.), 1938. *J. Sci. Instrum.* **15**, 5.
CRITTENDEN and TAYLOR (A. H.), 1929. *Trans. Illum. Engng. Soc., N.Y.*, **24**, 153.
DRESLER, 1937. *Licht*, **7**, 81.
DRESLER, 1938. *Handbuch d. Lichttechnik* (Berlin : Springer), 292.
DRESLER, 1940. *Licht*, **10**, 79.
DZIOBEK, 1934. *Licht*, **4**, 150.
DZIOBEK and REEB, 1933. *Z. tech. Phys.* **14**, 350.
FABRY, ROUX and PERRIN, 1929. *Rev. Opt.* (*theor. instrum.*), **8**, 1.
FORSYTHE, 1935. *Trans. Illum. Engng. Soc., N.Y.*, **31**, 181 (see Discussion, pp. 197–9).
FORSYTHE, BARNES, B. T. and SHRIDER, 1938. *J. Opt. Soc. Amer.* **28**, 241.
GERSHUN, 1932. *Phys. Z. Sowjet*, **2**, 149 (in German).
GERSHUN, 1936. *Svetouse pole* (*The Light Field*) (Moscow).
GIBSON, 1935. *J. Opt. Soc. Amer.* **25**, 131.
GIBSON, TEELE and KEEGAN, 1938. *J. Opt. Soc. Amer.* **28**, 178.
GIBSON, TEELE and KEEGAN, 1939. *J. Opt. Soc. Amer.* **29**, 144.
GIBSON and TYNDALL, 1923. *Sc. Papers Bur. Stds.* **19**, 131.
GUILD, 1932. *Reports of Discussion on Vision* (London : The Physical Society), 60.
HARRISON, H. T., 1929. *Proc. World Engng. Cong.* (Tokio), **24**, 181.
HECHT and WILLIAMS, 1923. *J. Gen. Phys.* **5**, 555.
HYDE, 1907. *Bull. Bur. Stds.* **3**, 81.
IVES, 1915 a. *Trans. Illum. Engng. Soc., N.Y.*, **10**, 235.
IVES, 1915 b. *Phys. Rev.* **6**, 334.
IVES, 1919. *J. Franklin Inst.* **188**, 217.
IVES, 1924 a. *J. Franklin Inst.* **197**, 147.
IVES, 1924 b. *Com. int. d'Eclair*, vi^{e} Sess. 77.
IVES and KINGSBURY, 1915. *Phys. Rev.* **6**, 319.
JAGGI, 1939. *Helv. Phys. Acta*, **11**, 432.
JAINSKI, 1938. *Diss. Tech. Hochschule, Berlin.*
KÖNIG, A., 1892. *Ann. Phys. u. Chem.* **45**, 604.
KÖNIG, H., 1934. *Helv. Phys. Acta*, **7**, 433.
KÖNIG, H., 1937. *Helv. Phys. Acta*, **10**, 165.
KÖNIG, H., 1938. *Helv. Phys. Acta*, **11**, 432.
KÖNIG, H., 1939. *Helv. Phys. Acta*, **12**, 229.
LAMBERT, 1760. *Photometria.*
LITTLE and ESTEY, 1935. *Trans. Illum. Engng. Soc., N.Y.*, **32**, 628.
LOMMELL, 1880. *Ann. Phys. u. Chem.* **10**, 449 and 631.
LOMMELL, 1889. *Ann. Phys. u. Chem.* **36**, 473.
MOON, 1935. *J. Franklin Inst.* **219**, 17.
MOON, 1940. *J. Opt. Soc. Amer.* **39**, 195.
MOON and MILLS, 1934. *Rev. Sci. Instrum.* **6**, 8.
MOON and STEINHARDT, 1938. *J. Opt. Soc. Amer.* **28**, 148.
MOON and TIMOSHENKO, 1939. *J. Math. Phys.* **18**, 51.
PATERSON, 1932. *Illum. Engr., Lond.*. **25**, 308.

PRESTON, 1936 a. *Proc. Phys. Soc.* **48**, 869.
PRESTON, 1936 b. *Instn. Elect. Engrs.* **79**, 424.
PRESTON, 1938. *Proc. Phys. Soc.* **50**, 398.
PRESTON, 1939. *Proc. Phys. Soc.* **51**, 757.
PRESTON and MCDERMOTT, 1934 a. *Proc. Phys. Soc.* **46**, 256.
PRESTON and MCDERMOTT, 1934 b. *Proc. Phys. Soc.* **46**, 25.
RIBAUD, 1933 a. *Rev. Opt.* (*theor. instrum.*), **12**, 289.
RIBAUD, 1933 b. *Com. int. Poids et Mes., Proc. verb.* **18**, 247.
SHARP and SMITH, H. A., 1928. *Trans. Illum. Engng. Soc., N.Y.*, **23**, 428.
STILES and CRAWFORD, 1933. *Proc. Roy. Soc.* B, 112.
TEELE, 1941 a. *Bur. Stand. J. Res., Wash.*, **27**, 217.
TEELE, 1941 b. *J. Opt. Soc. Amer.* **31**, 696.
TEELE and SCHUHMANN, 1939. *Bur. Stand. J. Res., Wash.*, **22**, 431.
URBANEK and FERENEZ, 1939. *Com. int. d'Eclair.* x^{e} Sess. (not yet published).
VOOGD and WENT, 1939. *Philips Techn. Rev.* **4**, 260.
WAIDNER and BURGESS, 1909. *Elect. World, N.Y.*, **52**, 625.
WALSH, 1933. *Illum. Engr., Lond.*, **26**, 64.
WALSH, 1940. *Light and Lighting*, **33**, 162.
WEAVER, 1937. *J. Opt. Soc. Amer.* **27**, 36.
WENZEL, ROESER, BARBROW and CALDWELL, 1931. *Bur. Stand. J. Res., Wash.*, **6**, 1103.
WENZEL, ROESER, BARBROW and CALDWELL, 1934. *Bur. Stand. J. Res., Wash.*, **13**, 160.
WHITMORE, 1939. *J. Math. Phys.* **17**, 218.
WINCH, 1930. *J. Instn. Elect. Engrs.* **68**, 533.
WINCH, 1940. *Trans. Illum. Engng. Soc., Lond.*, **5**, 93.
WINCH and HARPER, 1932. *G.E.C. Journal*, **3**, 223.
WINCH and PALMER, 1934. *Illum. Engr., Lond.*, **27**, 149.
WINCH, PALMER, MACHIN and DUDDING, 1934. *G.E.C. Journal*, **5**, 125.
YAMAUTI, 1926. *J. Opt. Soc. Amer.* **13**, 561.
YAMAUTI, 1931. *Proc. Int. Illum. Cong.* (*Cambridge*), **1**, 188.
YAMAUTI, 1932. *Res. Elect. Lab., Tokio*, no. 339.
YAMAUTI, 1937. *Rev. gen. Elect.* **42**, 293.

DISCHARGE PHENOMENA IN GASES

By R. WINSTANLEY LUNT,* A. von ENGEL† AND J. M. MEEK‡

§1. INTRODUCTION

For reasons not unconnected with the war, the present *Report* deals with only three sections of the field, and seeks to give a critical summary of the present trend of thought without attempting to refer to all the relevant contributions. It attempts to extend the consideration of these sections given in the reviews of Druyvesteyn and Penning (1940) and of Francis and Jenkins (1940).

§2 (by R. W. L.) deals with the statistical theory of reactions in gases maintained in a state of partial ionization, and with the implications of this theory to the general understanding of the mechanism of discharges for which the kinetics of the simple ionization reaction by electron impact are of particular importance.

§3 (by A. v. E.) is concerned with the mechanism of the negative zones in glow discharges, especially with the factors controlling the cathode fall of potential; in this the absolute value of the velocity coefficient of the simple ionization reaction is a matter of moment.

§4 (by J. M. M.) deals with the mechanism of the rapidly changing phases of partial ionization which culminate in a spark; and here again the same reaction is of prime importance in the interpretation of the observed phenomena.

§2. THE STATISTICAL THEORY OF DISCHARGE REACTIONS (R. W. L.)

(i) *Introduction*

This approach to the problems of discharge reactions has developed in recent years largely as an attempt to provide a basis for an understanding of the kinetics of chemical change in partially ionized gases. But it is now apparent that it concerns reactions of many kinds, some of which are of prime importance to the mechanism by which a discharge occurs and is maintained, and that there is often little purpose in preserving a distinction between "physical" and "chemical" changes. Ionization by electron impact was the reaction first considered from this point of view (Townsend, 1901; Kirkby, 1902; Townsend, 1903, 15); it was then recognized that the experimental data, which relate to

* Research Department, Messrs. Callender's Cable and Construction Co., Ltd.
† Electrical Laboratory, Oxford.
‡ Research Department, Messrs. Metropolitan Vickers Electrical Co., Ltd.

inhomogeneous electron beams passing through a gas in a uniform electric field, and, therefore, to a range of very low current density, could be conveniently discussed in terms of a reaction velocity coefficient, Townsend's α, the number of ionizing collisions per cm. of electron path in the direction of the field. Townsend and Kirkby then perceived that the theory should be based on the idea that there was a critical electron energy below which the reaction would not occur; but their simple theory lacked, amongst other things, any precise knowledge about the dependence of the probability of reaction on the electron energy. Kirkby (1911) foresaw that the ionization reaction should be the prototype of other reactions initiated by collisions between electrons and gas molecules not merely in the conditions in which the ionization reaction has been examined experimentally, but also for conditions in glow discharges in which the degree of partial ionization and the current density is at least six orders of magnitude greater. For one reaction involving chemical change he succeeded in demonstrating that, in the uniform positive column of a glow discharge, the kinetics of reaction can be described by a velocity coefficient formally similar to Townsend's α and having apparently similar characteristic properties.

(ii) *Recent theory*

The more recent theory attempts to trace the relationship between average quantities characteristic of reactions, such as reaction rate coefficients, to the data for individual collisions between the species present which engender reaction in partially ionized gases (Emeléus and Lunt, 1936; Emeléus, Lunt and Meek, 1936). Whilst the theory is applicable to collisions of any type between the various species present, we consider here only the particular form appropriate to reactions initiated by collisions between electrons and uncharged molecules of the gas because such reactions appear to be of predominant importance to an understanding of many discharge phenomena.

The theory's first achievement was to provide a detailed explanation for the observed characteristics of the velocity coefficient (α) of the simple ionization reaction; this depended on only one major and explicit assumption, a Maxwellian distribution of the electron energies, which is at least not improbable.

Mindful of Kirkby's demonstration of the apparent existence of a similar coefficient for reaction in the positive column, Lunt and Meek (1936) sought evidence for an extended validity of the theory and the auxiliary data to such conditions of greatly increased partial ionization and current density; in these there is independent experimental evidence that the assumption of an electron energy distribution differing but little from the Maxwellian is likely to be well founded. Whilst the conclusions still appear to be sound, it can now be seen that they were reached by an imperfect statement of the problems concerned involving implicit assumptions of which little account was taken. New experimental evidence was adduced to show that for at least one other reaction in positive-column conditions the characteristics of the velocity coefficient are explicable

in detail by the theory (Lunt and Meek, 1936; Lunt, Meek and Smith, 1937). The conditions under which the theory provides a general explanation of the main characteristics of the velocity coefficients of chemical change in the various zones of a glow discharge have been summarized by Emeléus and Lunt (1936).* These include the specification of a factor Φ which takes account of the stoichiometric relation between the primary products of an electron collision reaction and the ultimate species, if any, by which the reaction is observed in experiment, and also of the chance, f, that any of the species involved escapes destruction before reaching the measuring device. The factors controlling Φ have been examined in some detail by Lunt and Swindell (1940). Their analysis shows that a detailed knowledge of the generative and destructive processes is essential to the specification of f for any one species; and this involves the computation, for any given discharge conditions, of the absolute concentration of the species concerned, which, however, is conveniently considered in terms of the fractional concentration λ_s, the ratio of the concentration of the given species to the total concentration of all species present.

In this way the theory has drawn attention to the actual composition of a gas while it is in a state of partial ionization, and incidentally enables this to be roughly estimated. Much has been written on the effects exerted on the characteristics of a partially ionized gas by " impurities " present before ionization occurs. But comparatively little attention has been paid to the *a priori*, equally potent " impurities " which are inevitably generated, sometimes in considerable amount, and concomitantly with charged particles, from any gas when a discharge is maintained in it. The estimation of the λ_s's for the latter type of " impurities ", and the consideration of their relevance to the physical mechanism of discharges, is clearly a matter of fundamental importance to general discharge theory.

The success of the recent theory in providing detailed explanations for the principal characteristics of reaction velocity coefficients, as deduced from experiments on glow discharges, rests in part on the assumption that data drawn from electron beam experiments remain valid for current densities many orders of magnitude greater. In §3 and §4 of this *Report* are to be found examples, under a somewhat different guise, in which the experimentally determined velocity coefficient of the simple ionization reaction has been widely assumed to have, and in other cases appears to have, an enormous range of validity extending far beyond the range of current density to which the determinations relate. It is therefore pertinent to present the theory and its applications so as to illuminate this somewhat paradoxical position, the resolution of which is of importance to many aspects of the mechanism by which a discharge is initiated and maintained. A brief formal statement therefore precedes some illustrative applications, especially to steady-state uniform-field conditions which transpire to be those most amenable to simple consideration.

* For a case treated in detail, see Lunt, Pearson and Topley, 1940.

This formal statement relates to an ideal system which corresponds in many respects to the real systems of discharges. In applying the theory, correspondence between ideal and real systems has been assumed; but a complete analysis of the assumptions thereby involved is lacking, as also are the precise definitions of the variants of the ideal system appropriate to particular cases. Some relevant considerations are noticed later.

It will suffice to consider here two average quantities, and two derived quantities, in terms of statistical theory for a reaction attributable to the partial ionization of a gas through which electrons drift in a field X. In any volume element R, the rate per unit volume of any specified type of collision between electrons at the concentration n_e and "parent molecules" (i.e., molecules unchanged in consequence of the state of partial ionization) at the fractional concentration λ is given in a form convenient for many applications of present interest by Emeléus and Lunt (1936),

$$R = k\lambda p n_e \int_{V_0}^{\infty} V^{\frac{1}{2}} Q(V) f(V) dV, \qquad \ldots\ldots(1)$$

where k is a numerical constant depending on the units used, p is the total pressure reduced to standard temperature, $Q(V)$ denotes the probability cross-section for the particular type of collisions concerned for electrons of energy V and is zero when V lies below some critical value V_0, and $f(V)$ is the electron energy distribution function. A related quantity is T, the rate per unit volume at which electrons lose energy in effecting the given type of collision; this is given by

$$T = k\lambda p n_e \int_{V_0}^{\infty} V_L V^{\frac{1}{2}} Q(V) f(V) dV, \qquad \ldots\ldots(2)$$

where V_L denotes the energy lost in a single collision involving an electron of energy V (Lunt and Meek, 1936).

The velocity coefficient for the reaction associated with the specified type of collisions, formally similar to Townsend's α, is denoted by P, the activity of the discharge; i.e., the number of the specified type of collisions per unit (electron) current density or per unit (electronic) charge transported in the direction of X. Usually the motion of ions may be neglected compared with that of electrons, and P is given by

$$P = \frac{R}{n_e W}, \qquad \ldots\ldots(3)$$

where W denotes the drift velocity of the electrons. Caution must be used in applying this expression in the special case of reactions in which the primary products are charged. In the idealized case of a system isotropic in directions perpendicular to X (or the z-axis), β, the total activity of the partially ionized gas between two planes $z = 0$ and $z = z$ (i.e., the total number of collisions of the specified type occurring per unit (electronic) charge transported across the planes), follows as

$$\beta = \int_0^z P\, dz. \qquad \ldots\ldots(4)$$

Kirkby (1911) introduced the terms activity and total activity when considering chemical change.

To derive the corresponding quantities for an observed reaction product, which are distinguished by the suffix π, it is necessary to introduce the factor Φ, and we find

$$R_\pi = \Phi R \; ; \quad P_\pi = \Phi P \; ; \quad \beta_\pi = \int_0^z \Phi P \, dz. \qquad \ldots\ldots(5)$$

The factor Φ is conveniently regarded as the product of three terms, $\phi\psi f$ (Lunt and Swindell, 1940). Except for branching chain reactions, the stoichiometric factor $\phi\psi$ is a constant, specified by the initial electron collision type and by the observed product species, and is often a small number close to unity. The fractional utilization, f, depends on λ_g, the fractional concentration of the immediate product of the electron—parent molecule collisions which engender the formation of the observed product; λ_g may be found by solving the stationary state condition

$$\Sigma R_g = \Sigma R_d, \qquad \ldots\ldots(6)$$

where the R_g's and R_d's denote, respectively, the rates of possible generative and destructive processes (not necessarily involving electron collisions) for the species concerned, the R_d's being necessarily functions of λ_g. The fractional utilization then follows as

$$f = R_d'/\Sigma R_g, \qquad \ldots\ldots(7)$$

where R_d' denotes that part of ΣR_g associated with the observation and measurement of the product.

In general, it is to be anticipated that λ, f, and λ_g, for any particular electron collision reaction leading to a particular observed product, will depend on the factors governing the degree of partial ionization of the gas. But for many cases of practical importance, some of which are noticed later, it transpires that in a limited range of pressure and of low current density all λ_g's tend to be very small. It then follows that λ is sensibly constant and often nearly unity, that f is nearly unity, and hence that Φ is sensibly constant (except for branching chain reactions).

With this brief statement of the theory we may turn to consider some applications, and the extent to which they illuminate the possibilities and limitations of the theory.

If $Q(V)$, $f(V)$, and λ are independent of p and n_g, it follows from (1) and (3) that P/p is a function of X/p provided that W and $\overline{V}$, the mean value of the electron energy, are functions of X/p. For the simple ionization reaction in permanent molecular gases we thus find a likely explanation for the observed dependence of α/p on X/p, and also for the close agreement between the numerical values of α/p and P_i/p when the latter are computed from the (very low current density) data for W, $\overline{V}$, and $Q_i(V)$. Similarly, provided Φ is also constant, an explanation is at hand for the observation, for uniform positive-column conditions, that the experimental analogue of P_π/p for a reaction involving chemical

change (Kirkby, 1911) and for a physical reaction (Lunt, Meek and Smith, 1937) in each case depends on X/p, and in the latter case that $\Phi P/p$, as computed from theory, follows closely the trend of its experimental analogue. And the assumption of the invariance of λ, $Q(V)$, $f(V)$ and Φ, with p and n_e at all parts of the system suffices to account for a principal characteristic of chemical change (Emeléus and Lunt, 1936): that is, that $\bar{\beta}_m$ (the volume average reaction velocity coefficient, and formally similar to β) for reactions involving chemical change in glow discharges is often constant over considerable ranges of pressure and average current density.

The congruence between theory and experiment summarized in the previous paragraph is still remarkable despite the fact that in no single case has there been either a detailed theoretical attempt or an experimental one to justify *all* the various assumptions involved. This is especially the case for reactions in glow discharges; for then the congruence rests on the assumed applicability of data drawn from electron beam experiments to conditions in which the current density is at least six orders of magnitude larger. It is possible that there are mutually compensating variations with current density of the auxiliary quantities concerned in the computations of R and the dependent quantities P and β; and it has already been indicated that f, and hence ϕ, is likely to vary. The possibility of such variations is considered in the following detailed discussion of some particular applications of the theory to uniform field conditions in a steady state.

(iii) *Homogeneous electron beam experiments*

Such experiments provide data for $Q_i(V)$, the cross-sections for the simple ionization reaction $e+A \rightarrow A^+ + e + e$ (cf., e.g., Tate and Smith, 1932). The fact that the derived values of $Q_i(V)$ are not sensitive to pressure and current density for very low values of these two parameters justifies the attribution of the $Q_i(V)$ values to the single reaction given above; but it does not warrant the assumption that the same values will remain valid at higher pressures and current densities, when the possibility of interaction effects must increase. Alternatively, if the meaning of $Q_i(V)$ is slightly changed so that it refers to all collisions involving electrons which result in simple ionization of the parent molecules, it is at once apparent that there is no *a priori* reason for supposing that $Q_i(V)$ is independent of the concentrations of the two species concerned. Similarly, uncertainty surrounds any assumption of invariance of $Q(V)$ for other electron collisions, whether derived from experiment or from any theory which neglects the possibilities of interaction.

The published data for $Q_i(V)$ are, however, infected with a further uncertainty. They are stated to refer to a particular gas, often in a state of high "chemical" purity; but they refer necessarily to a mixture consisting of parent molecules and of species generated from them by electron impact. The procedure of Lunt and Swindell (1940) does not appear to have been used to estimate the

actual composition; and it is no more than a reasonable inference from the low current densities used in the experiments that the data relate to a gas consisting mainly of parent molecules.

(iv) *Inhomogeneous electron beam experiments*

Uniform field experiments of this type have been developed by the Townsend school (Townsend, 1915, 25) to measure W, $\overline{V}$, the mean value of $f(V)$, and α; the current density is known to be very low, but has rarely been recorded (Townsend, 1901). For a given variety of parent molecules W, $\overline{V}$, and α/p are often accepted as being functions of X/p only; but data supporting this contention are not very extensive. As in the case of $Q_i(V)$, it can be inferred from the low current densities involved that the data probably relate to gases consisting mainly of parent molecules. For W alone is there direct experimental evidence to show that the dependence of this quantity on X/p is substantially unchanged if the current density is increased by many orders of magnitude to uniform positive-column conditions (Harris, 1934; Lunt and Meek, 1936); and in these conditions it is much less certain that the gas consists mainly of parent molecules.

(v) *The mean fractional electron energy loss* χ

The published data are computed from the relation $\chi = 2{\cdot}46\, W^2/U^2$, where U is the mean velocity of random motion corresponding with the mean electron energy $\overline{V}$, and the numerical factor involves the assumption that $f(V)$ is Maxwellian in form (Townsend, 1925). Justification for this assumption is available for those gases for which it leads to numerical agreement between α/p and P_i/p as computed from (1) and (3) (cf. § 2 (vii), *infra*).

In the steady state, χ is the quotient of the total energy gained by electrons per unit volume and per unit time divided by ΣR_T, the corresponding total electron collision rate for all possible types of collisions; the summation must extend to all species present, R_T for any one species being given by (1) when λ relates to this species and $Q(V)$ denotes the total cross-section. Since each λ is likely to vary with the degree of partial ionization of the gas, so also is χ likely to vary; in addition, the published data are subject to all the limitations and uncertainties which attach to the data for W and $\overline{V}$. Ignoring all these complexities, von Engel and Steenbeck (1934) showed that the published data for χ can be used to compute positive-column characteristics for molecular gases in fair agreement with experiment. This remarkable result suggests that the low current data remain valid for these conditions; there are independent data then to justify the Maxwellian form for $f(V)$, but it is extremely improbable that the gas composition does not vary when the current density changes by a factor of 10^6 or more.

(vi) *The electron mean free path*

By definition the electron mean free path, Λ_e, is given by the quotient of $kn_e\int V^4 f(V)\,dV$, the total electron path per unit volume per unit time divided

by ΣR_T; for reasons already stated it cannot be expected to be independent of the current density. In the limiting case of very low current density, Λ_e, as given by this quotient, affords a detailed explanation of the connection between the free path of Ramsauer (cf. Brode, 1933) which depends on V, and the mean free path of Townsend (1925), which depends on $\overline{V}$ and X/p.

(vii) *The simple ionization reaction*

Some limitations and uncertainties surrounding the velocity coefficient α, as derived from experiment, have been stated in (iv); these should be qualified by Townsend's (1915) consideration of the soundness of the measurements and of their interpretation (cf., also, Loeb, 1939). It has already been mentioned that for some molecular gases the data for α/p are closely reproduced by computations of P_i/p from (1) and (3), using the experimental data for $Q_i(V)$, W, and $\overline{V}$, and the assumption that $f(V)$ is Maxwellian in form. For the experimental conditions appertaining to the measurement of α, the conclusion that $f(V)$ differs little from this form is qualified only by the uncertainties concerning the gas composition to which the auxiliary data relate. Historically, it is pertinent to note that an approximate formal expression for P_i/p was derived by von Engel and Steenbeck (1932) by a procedure which is equivalent to replacing $Q_i(V)$ in (1) by an approximation, and to using for W a theoretical expression involving χ and Λ_e as given by simple kinetic theory.

In any region of an actual discharge the rate of generation of electrons (and ions) can be described by a reaction velocity coefficient α_d, the number of electrons generated per cm. of electron path in the direction of the field. Whilst α_d is formally similar to α and P_i it is defined with respect to the product species; it thus takes account of all processes generating electrons, including the simple ionization reaction, which in some circumstances is convenient (cf. § 3 and § 4). How far the electron beam data for α afford even a rough guide to the value of α_d in any one zone of a glow discharge is a moot point. The present considerations anticipate that the sum of all the possible P_i terms will vary with the current density since the gas composition will vary, and hence that this contribution to α_d will cause α_d to depend on the current density and differ from α at any given values of X and p. It is, however, widely assumed that α_d does not differ significantly from α (cf. § 3 and § 4). There is little evidence for uniform field conditions. Killian's (1930) analysis of data for a uniform positive column of high current density shows that R_i as computed for the simple ionization reaction from (1) using electron beam data for $Q_i(V)$ and the assumption that $\lambda = 1$ accounts fairly well for the observed rate of electron destruction ; but the ionization of excited species, if present (i.e., $\lambda < 1$), could then lead to values of ΣR_i indistinguishable from those computed for R_i. Another case is provided by the theoretical derivation of the radial variation of the electron concentration in a uniform positive column (cf. Druyvesteyn, 1933). One assumption made in reaching agreement with experiment is equivalentto saying that $W_d \alpha_d$ is constant

irrespective of the current density $n_e W_d$ for given values of X and p, so that the rate of electron generation is $n_e W_d \alpha_d$, where W_d denotes the actual electron drift velocity. The rate computed from (1) and (3) is, however, $n_e W_d \Sigma P_i$, and is likely to reflect the radial variations in gas composition and current density. But provided W_d does not change much radially, an unrestricted validity of Townsend's α would justify the assumption made, for then $\alpha_d = \alpha$, and the rate of electron generation ($n_e W_d \alpha$) is proportional to n_e only. In the theory of the cathode dark space (§ 3) and of spark formation (§ 4) apparent agreement between calculation and experiment is sometimes reached by assuming that $\alpha_d = \alpha$; but again the significance of this is obscure because mutually compensating effects may intervene.

(viii) *The excitation of the hydrogen continuous spectrum*

This spectrum is associated with the transition $^3\Sigma_g^+ \rightarrow {}^3\Sigma_u^+$. For the ideal case of the gas consisting of ground-state molecules, the velocity coefficient P/p for the reaction generating $^3\Sigma_g^+$ molecules has been computed from (1) and (3) as a function of X/p, using electron beam data for W and $\overline{V}$, the Maxwellian form for $f(V)$, and theoretical estimates for $Q(V)$ (Lunt and Meek, 1936). Since Φ must be unity, the computed values of P/p were compared with the corresponding values of P_u/p derived from experiment for uniform positive-column conditions, P_u being in arbitrary units; the values of P/p follow closely those of P_u/p over a wide range of X/p (Lunt, Meek and Smith, 1939). This apparent and remarkable agreement between theory and experiment could arise if the electron beam data and the estimates for $Q/(V)$ have a greatly extended validity, and if in the experiments the gas consisted mainly of parent molecules, and the actual electron energy distribution differed little from the Maxwellian; but, as in previous cases, there are other possibilities. A nearly Maxwellian form for $f(V)$ is to be expected from the data for discharges in other gases; and hydrogen was rapidly streamed through the discharge in order to maintain λ not less than 0·9. To this extent the postulates of theory are roughly fulfilled, and hence the inference of an extended validity for W, $\overline{V}$, and $Q(V)$ is supported. But there are other uncertainties: it has yet to be shown that the modifications of the theory necessary to its application to a system with finite boundaries (the walls of the tube containing the positive column) and to streaming parent molecules are unimportant; and the experiments omitted a demonstration of the predicted current invariance of P_u.

(ix) *Kirkby's positive-column reaction*

Kirkby (1911) showed that $\overline{P}_m$, the average reaction velocity coefficient for the oxidation of hydrogen in electrolytic gas, varies so that P_u/p appears to be a function of X/p only. From this he deduced that the reaction kinetics must find their explanation in terms of collisions between electrons and parent molecules; he thus paved the way to the recent theory, although his postulated

mechanism and formal theory are no longer acceptable. As in the case of the two reactions just considered, the general character of Kirkby's data finds an explanation in terms of (1) and (3) and the assumption of the extended validity of the electron beam data; but no attempt has yet been made to show that the theory can lead to a detailed understanding of the reaction mechanism and its characteristics. It is noteworthy that the elegance and thoroughness of Kirkby's experimental procedure has rarely been surpassed.

(x) *Electron attachment reactions*

These have been examined mainly in inhomogeneous electron beam experiments (cf. (iv), *supra*). The velocity coefficient is often expressed in terms of the attachment constant h; the relation between h and the type of coefficient used here is that P_-, the experimentally derived counterpart of P for such reactions, is given by

$$P_-/p = 1{\cdot}32 \,.\, 10^{15}(X/p)W^{-\frac{1}{2}}h, \qquad \ldots\ldots(8)$$

and the data show that P_-/p depends on X/p (Bailey and Higgs, 1929). The statistical approach to such reactions is sometimes helpful in throwing light on the mechanism. In oxygen, for example, the rapid rise in P_-/p as V increases beyond about 1·6 ev. has been attributed to a reaction consequent on an electron collision reaction with parent molecules generating $^1\Sigma^+$ molecules, for which $V_0 = 1{\cdot}62$ ev. (Bradbury, 1933; Loeb, 1939). But for any likely form of $Q(V)$ for this latter reaction, and for any likely form of $f(V)$, it is readily shown from (1) and (3), using (legitimately) the electron beam data for W and $\bar{V}$, that the rapid rise in P/p occurs for values of $\bar{V}$ much below 1·6 ev. It follows that the postulated attachment mechanism is improbable.

(xi) *The electron energy balance*

For the steady state, equality between the gain of energy from the field and the loss by collisions requires from (2) that in any volume element

$$\Sigma T = n_e XW, \qquad \ldots\ldots(9)$$

where the summation extends to all possible electron collisions involving all other species present, each associated with a λ. For uniform field conditions Lunt and Meek (1936) showed that, if the gas consisted mainly of parent molecules, for the six principal reactions in hydrogen ΣT, computed for an assumed Maxwellian form for $f(V)$, amounted to about $0{\cdot}8\,n_e XW$ over a tenfold range of X/p; the calculations were based on electron beam data, and, in part, on theoretical estimates for the cross-sections. This result was considered to support the conclusion reached from the calculations of P_i/p, that the actual distribution differs but little from the Maxwellian. It can be seen that this conclusion has some justification for values of the current density, Wn_e, comparable with that in the electron beam experiments to which the auxiliary data relate.

At this point it should be noticed that, as in other cases previously considered, there has been no demonstration that expressions (1) and (2) derived for an ideal system are applicable to conditions in a discharge. Tonks (1936) pointed out that it is necessary to take into account the consequences of the real boundary to an actual discharge, and that ΣT in (9) should contain a term representing the energy conveyed to the boundary; when this correction is made for the practically important case of a cylindrical insulating boundary (uniform positive-column conditions), the numerical factor 0·8 is slightly increased without, however, changing materially the significance of the results, except that they now correspond to actual conditions in a discharge in this one respect. The electron energy loss at the boundary is, however, largely a consequence of the disappearance of electrons there, and is associated with a radial variation in concentration (cf. (vii)). This at once suggests the necessity of considering the destructive processes operative, and the radial variation in concentration, for all other species present. The gas concentration must also change progressively unless generative processes are counterbalanced by destructive ones. At any current density, therefore, the discussion of the steady state, supposedly relating to some particular gas, calls for the prescription of the generative and destructive processes for all species; and because the rates of these processes are likely to vary with the current density, it must be anticipated that the gas composition will change correspondingly.

Seen from this point of view the foregoing illustrative applications of elementary statistical theory do but indicate the kind of information which a more detailed and precise treatment is likely to provide.

In so far as the experiments referred to in these applications on the one hand and the main tenets of discharge theory on the other are related to each other through the assumption that the partially ionized gas consists mainly of parent molecules independently of the current density, a more precise consideration of reaction kinetics is a matter of considerable importance. And yet, while these applications are sullied by more or less plausible assumptions where experimental data are lacking, and the issue may be confused by unidentified mutually compensating effects, they constitute an apparent link between widely diverse data ranging from reaction kinetics of chemical change not involving partial ionization of the gas, through those associated with partial ionization, to the simple ionization reaction which is of paramount importance to the mechanism of all discharges.

§3. THE MECHANISM OF THE CATHODE FALL IN GLOW DISCHARGES (A. v. E.)

(i) *The simple theory of the cathode fall*

To appreciate the present position of the theory of the cathode fall in a direct current glow discharge it is convenient in the first place to make use of a simple model. This consists of a plane cathode of uniform material and high electrical

conductivity and a remotely situated anode, the inter-electrode space being filled with a single gas. The cathode is a source of "primary" electrons in consequence of positive ions impacting on it; in addition there is a source of positive ions attributable to ionizing collisions between electrons and gas molecules, these collisions increasing in frequency towards the boundary between the cathode dark space and the negative glow. Extending from the cathode there is an electric field which controls the generation, motion, and destruction, particularly at the cathode, of the charged particles which carry the current. At the cathode electrons are set free by ionic impacts, one electron, apart from those captured by ions, being set free for about 10 to 100 ion impacts; an electron set free in this manner has, on an average, an energy of about 1 ev. At the cathode surface the current is therefore transported mainly by positive ions and is sensibly determined by the product of the ionic velocity and concentration, since the relative concentration of electrons is small. The distribution of positive ions then specifies the electric field between the two sources of carriers. The ionizing collisions effected by fast electrons accelerated by the electric field in the dark space constitutes a source of positive ions and "secondary" electrons which cumulates in the large density of ionization recognized as the negative glow. At the boundary between the dark space and the negative glow the field is small and hence, on account of the relatively large mobility of the electrons, the current there is transported mainly by electrons. The cathode fall of potential is then the difference in potential between the cathode and a plane in or near the dark-space—negative-glow boundary such that small changes in position of this plane affect this difference in potential insignificantly.

This simple model neglects effects attributable to temperature, oscillations, recombination, unexcited and excited molecules, charge transfer, negative ions, photons, impurities, variations of gas pressure, and to the walls. The corresponding theory was developed by von Engel and Steenbeck (1934). They abandoned the assumption, made by Compton and Morse (1928), that the potential distribution in the dark space is such that ionization by electron impact is a maximum, and, in its place, assumed that X, the field in the dark space, varies linearly with z, the distance from the cathode (at which the maximum value occurs); this latter assumption is in accordance with the measurements of Aston (1911), Brose (1925), Frey (1932), Steubing (1932) and Ernst (1935). This theory also supposes that γ_e, the number of electrons liberated at the cathode per ion impact, is constant, and that v_+, the velocity of the ions, varies linearly with X, that is, that the mobility μ_+ at a given pressure is constant. If d denotes the distance from the cathode to the boundary between the dark space and the negative glow at which X becomes sensibly zero, then V_c, the cathode fall of potential, is given by $V_c = \int_0^d X\,dz$, and is thus insensitive to small changes in d. Using Poisson's equation and the condition for continuity of the current, the

theory leads to the following relation between V_c, d, the thickness of the dark space, and I, the total current density,

$$I = \frac{V_c^2}{d^3}\frac{\mu_+}{\pi}(1+\gamma_c). \qquad \ldots\ldots(1)$$

The steady-state requirement is that N, the number of ionizing collisions in the dark space per "primary" electron, is such that the positive ions generated liberate, on reaching the cathode, one more "primary" electron; i.e.,

$$N = \int_0^d \alpha_c dz = \log(1+1/\gamma_c), \qquad \ldots\ldots(2)$$

where α_c denotes the number of ionizing collisions per cm. of electron path through the dark space and in the direction of X. The theory proceeds by assuming that α_c can be replaced by Townsend's ionization coefficient α (cf. § 2), which is supposed to be given by the semi-empirical expression $\alpha = pAe^{-pB/X}$ in which p denotes the gas pressure and A and B are constants specific to the gas; after some re-arrangement it then follows, from (1) and (2), that the relation between the cathode fall, the pressure, and the current density can be expressed by

$$(C_2 I/p^2)^{2/3} = (C_1 V_c)^{1/3} f(C_1 V_c C_2 I/p^2)^{1/3}, \qquad \ldots\ldots(3)$$

where $f(x) = \int_0^x e^{-1/y}\,dy$ and C_1 and C_2 are constants involving A, B, $\mu_+ p$ and γ_c.*

The notation of (3) corresponds with the principle of similarity according to which I/p^2, $\mu_+ p$, α/p, X/p, pd etc. are the basic variables.

Despite its complicated form, relation (3) is remarkable in that it not merely has a form independent of the gas and the cathode material (figure 1), but, as will be seen shortly, it represents experiment over a wide range of conditions. For any given value of I/p^2 there is a minimum for V_c according to (3) which is to be identified with V_n, the cathode fall of potential in the "normal" conditions (Bär, 1927), the corresponding values of I and d being denoted by I_n and d_n, respectively. From (1) and (3) it is possible to derive the dependence of d on I, V_c, and X_c, the value of X at the cathode. The calculated dependence of V_c/V_n, d/d_n and X_c/X_n on I/I_n is shown in figure 1. These results proved to be substantially consistent with the linear field in the dark space.

(ii) *Comparison between the simple theory and experiment*

The measure of agreement obtained between the foregoing simple theory and experiment is illustrated in figures 2, 3 and 4. Figures 2 and 3 show V_c/V_n as a function of I/I_n and d/d_n as a function of V_c/V_n for an iron cathode in various rare and molecular gases; the experimental curves are derived from the data obtained by Güntherschulze (1928, 30a); the ratios, as computed from the simple theory, are represented by dash-dotted lines. The more recent results obtained by Druyvesteyn (1938) for a carbon cathode in rare gases are shown in figure 4; Druyvesteyn examined these experimental curves in the light of the

* The definite integral results from using the linear field $X = X_c(1-x/d)$ and substituting, in (2), $y = (X_c/Bp)(1-xd)$. The upper limit becomes $x = X_c/Bp \propto V_n/d \propto (V_c I)^{1/3}$ when d is eliminated by means of (1). The value of the integral may be obtained from tables.

simple theory and assumed that these curves intercept at the ordinate $V_c/V_n = 2$ in order to avoid any reference to measurements of I_n. It is noteworthy that these results confirm the principle of similarity over a large range of I/p^2.

Reliable experimental evidence regarding the variation of the electric field strength is still lacking: field measurements mentioned in § 3 (i) show that the theory leads to the right order of magnitude for X_c.

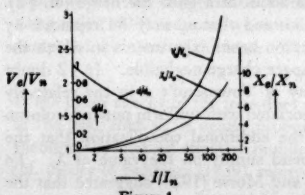


Figure 1.

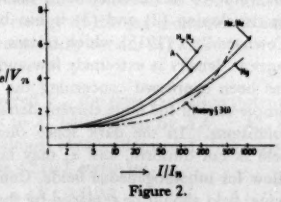


Figure 2.

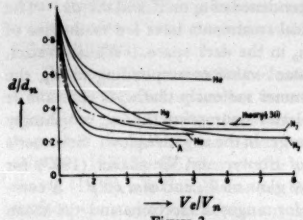


Figure 3.

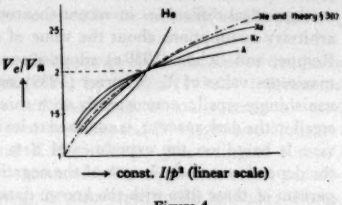


Figure 4.

Figure 1. Cathode fall, cathodic field strength and thickness of cathode dark space as a function of current density in terms of "normal" values according to theory, § 3 (i) (von Engel–Steenbeck).

$$\left.\begin{matrix} V_c/V_n \\ X_c/X_n \\ d/d_n \end{matrix}\right\} = 1 \text{ for } I/I_n \leqslant 1.$$

Figure 2. Cathode fall at iron cathode as a function of the current density from Güntherschulze's experiments (1928, 30 a).

Figure 3. Approximate thickness of cathode dark space at iron cathode (Ne, A at Mg–Al alloy cathode) as a function ofcathode fall, in terms of "normal" values (Güntherschulze,1928, 30 a).

Figure 4. Cathode fall at graphite cathode as a function of current density (Druyvesteyn, 1938).

It is apparent that theory is largely in accordance with experiment; there are, however, marked deviations for helium and smaller ones for argon. While these deviations must be attributed in part to experimental errors and, in part, to the simplifying assumptions in the theory, they may also arise from the possibility, which has been suggested (Druyvesteyn, 1938), that in the case of helium the velocity of the ions is not proportional to X, as has been assumed, because of the formation of He_2 molecules, and also that in the dark space the relation

between the random and drift velocities of the electrons departs from that in the experiments to measure α so that α_c/p does not depend on X/p only—which is also contrary to the assumption implicit in (3).

(iii) *Defects and omissions of the simple theory*

(a) *The coefficient of ionization, α_c.* Considerable uncertainty exists about the value of this coefficient in the various zones of a glow discharge (cf. §2). In developing (2) and (3) it has been assumed that α_c may be replaced by Townsend's α (1915), which relates to electron beam experiments in which the current density is extremely low and the space charge negligible. In §2 doubt has been expressed concerning the validity of Townsend's α in the relatively simple case of the large current density associated with uniform positive-column conditions. In the dark space there is the additional complication that the field is not uniform, and α_c may not depend simply on the value of X. To allow for inhomogeneous fields, Compton and Morse (1928) suggested that the actual field should be replaced by that corresponding to a plane half a mean free path nearer the cathode; this leads to a dependence of α_c on X and dX/dz. The mathematical difficulties in recent theoretical treatments have led to the use of arbitrary assumptions about the value of α_c in the dark space. Whilst Weizel, Rompe, and Schön (1939a) adopt a constant value corresponding to X_c, the maximum value of X, Scherzer (1939) assumes variously that α_c is constant or vanishingly small; concurrently with this latter assumption (i.e., α_c vanishingly small in the dark space) α_c is assumed to be larger in the negative glow. Scherzer's view is based on the experimental data of Brewer and Westhaver (1937) for the dependence of the length of the negative glow on V_c and of d on p. A comparison of these data with the known data for ranges of electrons and the mean electron energy per ionizing collision leads to an estimate of from 0·5 to 1 for the number of positive ions formed per "primary" electron. Such a small number corresponds to a small value of α_c and demands a very high value of γ_c, for which, however, no acceptable explanation is offered. The experimental results of Brewer and Westhaver are also inconsistent with the principle of similarity.

Apart from the uncertainty in d, there is the possibility that the lower limit of the integral in (2) should not be taken as zero but as Δ, the somewhat ill-defined distance from the cathode at which ionizing collisions first occur (Druyvesteyn, 1938). Other suggestions (Druyvesteyn and Penning, 1940) are that the appropriate value of α_c is that given by Townsend's α, corresponding to the mean field in the dark space, and that N should be computed from the data of electron beam experiments.

(b) *The velocity of the positive ions, v_+.* In deriving (1) it has been assumed that $v_+ = X\mu_+$, where μ_+ is the Langevin mobility and is given by

$$\mu_+ \sim e\Lambda/(M_+\bar{v}), \qquad \ldots\ldots(4)$$

where Λ denotes the mean free path of ions of mass M_+ and r.m.s. velocity v, which are assumed to arise from simple ionization of the gas and to be in thermodynamic equilibrium with the gas molecules (Tyndall, 1938). When the electric field is large, theory leads to the expectation that v_+ is no longer given by $\mu_+ X$ but (see Francis and Jenkins, 1940, p. 254, equation (16), when $\Omega \ll X$) by

$$v_+ = (2e\Lambda/\pi M_+)^{1/2} X^{1/2} = k_+ X^{1/2}, \qquad \ldots\ldots(5)$$

which is also the approximate form to which the expression of Compton and Langmuir (1930) reduces, when the mean energy of the gas molecules is small compared with that of the ions. The corresponding variant of (1) is

$$I \propto V_e^{3/2} d^{-5/2}. \qquad \ldots\ldots(6)$$

If the mean energy of the ions does not exceed V_a, the least excitation potential of the gas, then v_+ is nearly proportional to X, being given by $X k'_+$; k'_+, the theoretically derived effective mobility, is almost independent of X, being given (cf. theory of Scherzer (1939), referred to in § 3 (iv), *infra*) by

$$k'_+ = e\Lambda/[3{\cdot}5\, M_+(V_a + eX\Lambda/14)]^{1/2}. \qquad \ldots\ldots(7)$$

Apart from these theoretical estimates, experiment shows that when X/p exceeds a certain critical value there is a sudden increase in the mobility, and hence in v_+ (Tyndall, 1938).

Evidence that ions suffer inelastic collisions with the molecules of the gas in traversing the dark space has been discussed by Gehrts (1936). Information about the ion velocity distribution and the ratio of ion current to total current at the cathode has been provided by Campan (1934) (see also Güntherschulze, 1930 b; Güntherschulze and Keller, 1931). According to Runge's theory (1930), which neglects the ion accommodation coefficient at the cathode surface, the major part of the ion energy is transferred to the cathode; Runge estimates the amount of this energy to be from 94 to 55 per cent, according as the number of ion collisions, in traversing the dark space, is supposed to vary from 6 to 30; i.e. the energy lost in these collisions is from 6 to 45 per cent. The occurrence of fast-moving ions which suffer charge transfer (Kallmann and Rosen, 1930; Wolf, 1937; Francis and Jenkins, 1940, p. 257) in the dark space, thereby generating fast neutral particles, may diminish the effective mobility and influence the gas flow and pressure there, and thus change the electrical properties (see Mott and Massey, 1933).

(c) ***The electron emission factor, γ_e.*** As in the case of the coefficient α_e, it is convenient to distinguish between γ, the value of this factor as determined by ion beam experiments, and γ_e, which is the value characteristic of a cathode surface in a glow discharge, and, although still defined per ion impact, takes account of the electron emission attributable to other processes occurring simultaneously, for example photon impact. Photons are formed in the negative glow and in the dark space, and may reach the cathode surface. Thomson and Thomson (1933, vol. **2**, p. 296) state that if a plane metal cathode is replaced by one of gauze there

is no change in the discharge current, and suggest that the photon contribution to γ_c is unimportant. [The extent of ionization of gas molecules by ion impact is still uncertain (Townsend, 1939).] An increase of γ_c with ion velocity has often been deduced from discharge data (see §3(iv), Weizel *et al.*); such an increase is to be expected from ion beam experiments (see, e.g., von Engel and Steenbeck, 1932, p. 116). The nature of the gas layer adsorbed on the cathode markedly affects the value of γ_c (Oliphant, 1930; Paetow and Walche, 1938).* Metastable species which may be impelled by positive ions liberate electrons at a cathode surface. At a metal surface positive ions are sometimes reflected as metastable atoms accompanied by the emission of an electron (Oliphant, 1929; Gerthsen, 1928). It has not yet been established whether γ_c depends on the current density; such a dependence would be relevant to the consideration of cumulative ionization (Rogowski, 1939; Schade, 1938). An attempt has been made by Fucks (1933) to derive the equilibrium current density at the cathode which is influenced mainly by the radial component of the electric field and less by diffusion. Experiment shows that as V_c is increased, the current distribution at the cathode becomes progressively less uniform (Chiplonkar, 1940). Other phenomena affecting γ_c are that V_n for a single crystal cathode depends on the crystallographic plane acting as the cathode surface (Kurzke, 1936), and that V_n may increase or decrease at the melting point of the cathode, depending on the material (Wehnelt and Kurzke, 1934).

(d) *The influence of the negative glow on the dark space.* It has been estimated for the "normal" condition that the number of positive ions generated in the dark space is of the same order as that in the negative glow, and that the latter number becomes relatively larger as the conditions become "abnormal" (Druyvesteyn and Penning, 1940). In the negative glow, ions may be formed by photon impacts and collisions of the second kind (Kenty, 1933) in addition to electron impacts with gas molecules. Because the electric field in the negative glow is weak, this zone can be regarded as a plasma which is a large source of electrons and slowly moving ions (Langmuir and Tonks, 1929; Emeléus and Harris, 1927; Emeléus, Brown and Cowan, 1934; Seeliger, 1934). It is not known precisely, however, how the ions disappear. Experiments in which the negative glow was progressively restricted have led to the suggestion that recombination on the wall is negligible compared with that in the gas (Fischer, 1936: see also Druyvesteyn, 1930; Druyvesteyn and Warmoltz, 1931; Sloane and Emeléus, 1931). It seems probable, therefore, that the properties of the

* γ_c has been found to depend on X/p (see Francis and Jenkins, 1940); in rare gases, as the mass of the atom increases, the minimum value of γ_c occurs at higher values of X/p. Pike (1934) expresses doubts about the possibility of defining V_n experimentally because he sometimes observed large variations of V_n in very pure rare gases when using a cathode consisting of a metallic layer evaporated on to a ground glass surface. The variation may have arisen either from gas adsorbed on to the cathode during previous experiments (Spiwak and Reichrudel, 1936) or from the slow contamination of the metallic layer by a constituent of the glass foundation—both affecting γ_c.

dark space should no longer be regarded as independent of those in the glow; this has been illustrated by the use of "hollow cathodes", a case which has been discussed by von Engel and Steenbeck (1934) and by van Voorhis and Shenstone (1941). To allow for I_b the initial ion current into the dark space at the boundary between this zone and the negative glow (cf. § 3 (iv)) the logarithmic term in (2) should be replaced by $\log[1+(1/\gamma_c)(1-I_b/I)]$.

(e) *Interaction effects.* Interaction may contribute to the existing ionization. If radiation from an intense discharge through neon is passed into the dark space of another discharge, in either neon or neon containing traces of argon, a slight increase in the cathode fall is observed (Spiwak and Reichrudel, 1936). From variations in the current when the various negative zones are thus irradiated, Spiwak and Reichrudel conclude that the concentration of metastable species has a maximum towards the positive end of the negative glow; this deduction is in accordance with the spectroscopic observations of Emeléus and Duffendack (1935). When the current density and pressure are high, stepwise ionization may occur. There is evidence that the temperature of the gas in the dark space can attain values up to a few thousand degrees (von Engel, Seeliger and Steenbeck, 1933); in such conditions collisions between two neutral molecules may result in ionization. The influence of the gas temperature on V_e has been considered by Fischer (1938), who, from data for constant pressure, computed V_e as a function of I and d as a function of V_e for constant gas density (see Güntherschulze, 1928) using a calculated value of the mean temperature in the dark space.

(f) *The influence of negative ions.* Negative ions, on account of their small mobility, strongly affect space-charge distribution, but contribute only slightly to the transport of current (Massey, 1938). They are most readily generated in electronegative gases, at small electron velocities corresponding to 0·1 to 1 ev. and in the presence of excited species. Negative ions are also formed by positive ions striking a cathode and thereby extracting two electrons. This process (see Smith, 1938) has been observed in mercury (Arnot and Milligan, 1936), in molecular gases (Arnot, 1937), but not in argon (Timoshenko, 1941). The equilibrium concentration of negative ions in the negative zones of a glow discharge in neon, containing traces of chlorine, has been investigated by Emeléus and Sayers (1938) both theoretically and experimentally.

(iv) *Recent theoretical developments*

Several attempts have been made recently to improve the theory of the cathode fall of potential. The main achievement of these theories is to shed further light on the factors involved. Direct comparison with experiment is not practicable except in connection with some subsidiary phenomena: this is attributable, in part, to the intractable nature of the formal expressions which have been developed, and partly to the fact that these theories call for experimental data of a kind which has, for the most part, not been sought heretofore.

Weizel, Rompe, and Schön (1939a), extending Rogowski's theory of electrical breakdown (1932), start from the general equations given in § 3 (i). Using only (5), and assuming α_c to be given by the value of Townsend's α corresponding to X_c, the maximum value of the field, they derive the differential equation :

$$\frac{1}{\alpha_c}\frac{d}{dz}\left[X^{1/3}\frac{dX}{dz}\right] = -X^{1/3}\frac{dX}{dz} + \frac{4\pi}{k_+}I. \qquad \ldots\ldots(8)$$

A solution has been obtained assuming α_c=const.* The boundary conditions at the cathode and the boundary between dark space and negative glow are, respectively,

$$z=0 : I_-/I_+=\gamma_c \quad \text{and} \quad z=-d : X=0 : I_+=I_b ;$$

these yield with the substitutions

$$y=\frac{k_+}{6\pi I}\frac{dX^{3/2}}{dz}, \qquad y_0=\frac{1}{1+\gamma_c}, \qquad y_d=\frac{I_b}{I},$$

$$V_c=\left(\frac{6\pi I/p^3}{k_+p^{1/2}}\right)^{2/3}\left(\frac{\alpha_c}{p}\right)^{-5/3}\int_{y_d}^{y_0}\frac{\{y-y_d+\log(1-y)/(1-y_d)\}^{2/3}}{1-y}\,dy, \qquad \ldots\ldots(9)$$

$$pd=\frac{1}{\alpha_c/p}\log(1-I_b/I)/\gamma_c. \qquad \ldots\ldots(10)$$

Although the cathode fall cannot be computed simply from (9), this expression, together with (10), reveals the connection between k_+, α_c, and γ_c. Expression (10) contains the cathode fall implicitly because the value adopted for α_c depends on X_c/p, which latter reflects the value of V_c. According to (10), d becomes smaller as I_b/I increases. In the absence of precise data for I_b/I the theory cannot be exhaustively tested; when the few data available for I_b/I (Weizel and Fischer, 1935; Geiger, 1937) are used in (10), γ_c is found to increase, in agreement with experiment, from 0·02 to 0·06 as the ion energy at the cathode rises from 50 to 200 ev. The theory has been elaborated to permit a more detailed discussion of I_b/I (Weizel, Rompe and Schön, 1939b).

A few decades have passed since the last attempt to develop a general theory of the glow discharge. Scherzer (1939) has contributed an interesting one-dimensional analysis. This supposes that the negative glow provides the necessary positive ions for the dark space, and that recombination of ions and electrons in the gas balances their rate of generation (see, e.g., Loeb, 1939). The following additional assumptions are made : v_+ is given by k'_+X when k'_+ is given by (7); the space charge in the dark space is specified by the distribution of positive ions; the electron concentration n_e follows the Boltzmann distribution, $n_e=ae^{-eV/kT_e}$, where V denotes the space potential, T_e is the equivalent electron temperature, and a is a constant dependent upon the electron con-

* This assumption restricts the results to discharges with relatively high values of V_c since, when X_c is large, a change in X_c has only a small effect on α which is approximately equal to α_{max}.

centration in the plasma of the negative glow. The dependence of the positive-ion current on position is specified by α_c, which, together with the recombination coefficient ρ, is constant throughout the system; the consequences, however, of assuming α_c to be vanishingly in the dark space and large in the negative glow are also considered. Fast electrons, crossing the dark space without ionizing, carry a current I_s. Positive ions are generated mainly in the negative glow, in which zone is the anode. From Poisson's equation Scherzer then derives a third-order differential equation of potential which, omitting dimensional constants, reads

$$-\frac{k'_+}{\rho}\frac{d^2}{dz^2}\left[\frac{1}{2a}\left(\frac{dV}{dz}\right)^2+\frac{kT_e}{e}e^{\frac{eV}{kT_e}}\right]=e^{\frac{eV}{kT_e}}\frac{d^2V}{dz^2}+e^{\frac{2eV}{kT_e}}+\frac{\alpha_c}{\rho a}I_s, \quad \ldots\ldots(11)$$

At the anode boundary it is supposed that the current is attributable to I_s, to the diffusion of slow electrons from the plasma of the negative glow, and to the motion of positive ions.

No exact explicit solution of (11) is known. By approximation, and by dividing the system into dark-space, ionization, and recombination zones, relations were derived for each zone. Of the many results obtained, the following can be compared with experiment. For a discharge "unobstructed" in the axial direction, the predicted dependence of the electrode potential on the current is that these quantities are proportional when the distance from the anode to cathode is constant. As this potential is increased the dark space finally extends to the anode and the current diminishes. These predictions are in accordance with anomalies observed by Penning (1932). The predicted dark-space relation is

$$d \propto V_d^{1/3}/\alpha_c^2, \quad \ldots\ldots(12)$$

where V_d denotes the electrode potential and α_c now denotes the effective value in the plasma.

Seeliger (1940) pointed out that, since to a first approximation the negative glow can be treated as a "linear" plasma, it is permissible to discuss the motion of the carriers as a diffusion process (Seeliger, 1932) governed by the relation

$$D_a\frac{d^2n}{dz^2}=-c_1n+c_2n^2-c_3\frac{I_s}{n}, \quad \ldots\ldots(13)$$

where D_a is the ambipolar diffusion coefficient and the terms on the right-hand side of the equation correspond, respectively, to generation by random electrons, volume recombination, and generation by fast electrons energizing the glow. Secondary electrons produced by fast electrons are ignored. He showed that a connection exists between the two theories previously discussed: that while Scherzer introduces the Boltzmann distribution explicitly, Weizel, Rompe and Schön (1939 b) use a differential equation which takes account of the simultaneous occurrence of field and diffusional motion, which, for an indefinitely small drift current, is equivalent to the explicit introduction of the Boltzmann distribution.

§4. RECENT DEVELOPMENTS IN THE THEORY OF THE SPARK DISCHARGE (J. M. M.)

The coefficient of ionization effected by electrons moving largely at random through a gas but in an electric field is of fundamental importance in the consideration of the mechanism of the spark discharge. The value of this coefficient for the conditions in which breakdown occurs is denoted by α_b to distinguish it from the Townsend α (cf. §2), which relates to specific conditions, though in spark theory it is generally supposed that these two quantities do not differ significantly. It happens that an exact knowledge of α_b is not essential to one of the principal applications of spark theory, the calculation of the critical field for sparking. Such calculations are insensitive to the value adopted for α_b because of the exponential character of the equations involved. Thus it follows that the degree of agreement between the calculated breakdown potentials and those measured is not a reliable guide to the validity of any theory, and the exact picture of the mechanism as given by the theory must also be taken into consideration and viewed in its relation to the various observed sparking phenomena.

Ionization by electrons in a gas occurs at a field strength considerably below that necessary to cause breakdown, and many suggestions have been made about the conditions governing the increase in electron concentration requisite to cause a spark. Townsend's original suggestion (Townsend, 1915) was that a spark occurred at a field strength such that the positive ions also acquired sufficient energy to ionize the gas molecules; such ionization, however, must be excluded from further consideration because of its very low probability for the ionic velocities in the range of field strength concerned (Mott and Massey, 1933; Loeb, 1939; Varney, Loeb and Haseltine, 1940). Various other criteria have since been invoked (Holst and Oosterhuis, 1923; Taylor, 1927; Rogowski, 1928; Loeb, 1928, 39; Franck and von Hippel, 1929; Schumann, 1930; Holm, 1932; Sämmer, 1933; Thomson and Thomson, 1933; Kapzov, 1934; Schade, 1939); these include the critical conditions for the liberation of secondary electrons from the cathode by positive ion or photon impact. The criterion generally adopted for breakdown to occur is that more than one electron must be generated at the cathode by the positive ions which are formed in the passage of a single electron across the gap; since the process is cumulative, there is a resultant build-up of space-charge in the gap which leads to eventual breakdown. The formal expression of this criterion for a uniform field gap of length d cm. is that

$$\gamma_b \exp(\alpha_b d) \gtrsim 1, \qquad \ldots\ldots(1)$$

where γ_b denotes the number of electrons created at the cathode per positive-ion impact (Loeb, 1939) and the suffix b is used to denote the effective value of the electron emission factor for breakdown conditions (cf. §3 (iii) (c)).

The dependence of the electron generation coefficients α/p and γ on X/p, the ratio of the field strength to the gas pressure, is known for a limited range

of very low current densities and of low gas pressure, from electron beam experiments (cf. §2 and §3). If such data are used to calculate the minimum field-strength to cause a spark, using the criterion in the modified form

$$\gamma \exp(\alpha d) = 1, \qquad \ldots\ldots (1\,a)$$

good agreement is found between calculation and experiment for short gaps at low pressures. However, in the case of gaps of length greater than 0·3 cm. at atmospheric, or higher, pressure, the observed critical field is virtually independent of the cathode material (Loeb, 1939), and, moreover, is much below that at which a significant value of γ_b is to be anticipated from the data for γ. This indicates that for these latter conditions, which include a range of great practical importance, the mechanism of the spark depends on processes in the gas only; this inference is independent of any uncertainties about the relation between α and α_b. Furthermore, these processes must occur in the small time interval in which a spark develops, which the data show to be that corresponding with the transit time of an electron across the gap; in such small intervals the movement of positive ions is too small to be significant.

It has thus become apparent during recent years that the theory of the mechanism of spark discharge for values of pressure times gap length in excess of 200 mm. Hg. × cm. in air must find its explanation mainly in terms of ionization by electrons and photon impact with the gas molecules, together with the resultant distortion of the field by space charge. Recently a theory based on such considerations has been developed which explains the principal features of sparking and, when the initial field distribution is known, also leads to calculated values of the critical breakdown field in good agreement with experiment (Meek, 1940; Loeb and Meek, 1940, 41).

The new theory begins with the consideration of the development of a single "electron avalanche" initiated by an electron liberated at the cathode. In a uniform field (parallel plate) gap such an electron creates new electrons at the rate α_b per cm. advance in the field direction and the total number formed when the distance travelled is x cm. is $\exp(\alpha_b x)$, an equal number of positive ions being left behind. Such a progressive ionization of the gas in a more or less limited region and in the direction of the field is termed an avalanche; and because of their relatively large mobility, the electrons tend to congregate in the advancing head of the avalanche. When the head reaches the anode the electrons are absorbed in the metal, while the positive ions remain in a roughly conical channel whose base is at the anode. Because of the low relative mobility of the ions and their space distribution in the avalanche, the latter forms but a poorly conducting filament across the gap; that is, the development and transit of an avalanche across the gap does not of itself constitute a breakdown of the gap. Nevertheless, these nearly stationary ions set up a radial field (i.e., perpendicular to the inter-electrode field, and additional to the distortion in that direction) and contribute largely to the sparking mechanism. For when

the radial field attains a value of the order of the inter-electrode field it will attract towards the stem of the original avalanche any secondary avalanches initiated by neighbouring photo-electrons and which are progressing under the action of the inter-electrode field. Because of the exponential variation of the rate of ion generation with distance, the bulk of the positive ions are generated in the region close to the anode; there the radial field attains its maximum value, and hence there, too, the maximum directive effect is developed. As the heads of auxiliary avalanches are drawn into the primary avalanche the positive ions left behind extend the region of positive space charge. A second, and important, consequence is that the region behind the head of the main avalanche then consists not mainly of positive ions but of a densely ionized gas, containing both ions and electrons, which is therefore highly conducting; this phase, termed positive streamer formation, then develops backwards towards the cathode (Raether, 1939; Loeb and Meek, 1941). When the streamer reaches the cathode a conducting filament exists across the gap which constitutes the complete breakdown preceding the final arc. It follows that the new criterion for breakdown carries the requirement that the radial field is of the same order of magnitude as the inter-electrode field, namely, that

$$X_r \sim X, \qquad \ldots\ldots(2)$$

where X_r denotes the radial field strength.

The form of an electron avalanche has been observed by Raether (1939) in cloud-chamber experiments. He found that the rate of lateral diffusion is equal to that anticipated from theory, namely, that $\bar{r}$, the average radius of an avalanche after it has travelled a distance x, is given by $\bar{r} = \sqrt{2Dt}$, where $t = x/W$, and D is the coefficient of electron diffusion estimated from the electron drift velocity W.

The approximate value of the radial field can then be calculated (Meek, 1940), and is given by

$$X_r \sim 5 \times 10^{-7} \left(\frac{x}{p}\right)^{-\frac{1}{2}} \alpha_b \exp(\alpha_b x). \qquad \ldots\ldots(3)$$

Adoption of the criterion $X_r = X$ yields the equation

$$pd(\alpha_b/p) + \ln(\alpha_b/p) = 14{\cdot}5 + \ln(X/p) + 0{\cdot}5 \ln(d/p), \qquad \ldots\ldots(4)$$

which is a form convenient for solution. In this expression X can be regarded as computable in terms of the gap parameters p and d provided that the dependence of α_b/p on X/p is known independently. In the absence of data for this dependence, it is customary to assume that it may be taken to be the same as that of α/p on X/p as given by electron beam data (Townsend, 1915; Loeb, 1936). There is probably little uncertainty about the validity of this assumption in the regions of initial development of electron avalanches where the current density does not significantly exceed that in the measurements to measure α/p; but in the regions of more intense ionization, which are likely to be accompanied by appreciable

changes in the density and composition of the gas, the continued validity of α/p is very uncertain. Nevertheless the assumed validity of α for α_r in (4) leads to remarkably good agreement between calculation and experiment for the critical breakdown field for the range of high values of the product of the gap parameters pd, namely, for the range in which the simpler criterion in (1) is inadequate.

The insensitivity of the calculations to a more precise formulation of the criterion in (4) is illustrated by the following example. For $d=1$ cm. and $p=760$ mm., and for $X=X_r$ and $X=0{\cdot}1\,X_r$, the corresponding values for the critical value of X according to (4) are 32·2 kv.cm^{-1} and 31·6 kv.cm^{-1}, respectively, a difference of only 2 per cent. Differences of no greater order are found if the value assumed for α is changed by 10 per cent. The cause of this insensitivity is readily seen from an inspection of (4).

The agreement between the calculated and the measured breakdown potentials has been found to hold for values of pressure times gap length up to as high as 10,000 mm. Hg. × cm. There is, however, a slight deviation from Paschen's law, and the calculated breakdown potential at high pressure is lower than that for the same value of pressure times gap length at low pressure. Thus for a 1 cm. gap in air at 760 mm. Hg. the calculated breakdown potential is 32·2 kv., but is 31·5 kv. for a 0·1 cm. gap at 7600 mm. Hg. Recent experiments confirm the existence of such a deviation from Paschen's law (Skilling and Brenner, 1941; Trump, Safford and Cloud, 1941).

If the applied voltage is in excess of the minimum to cause breakdown, it is to be expected from the theory that an avalanche, originating at the cathode of a uniform field gap, will acquire the requisite ion density to fulfil the condition $X_r \sim X$, and hence develop into a streamer, before it reaches the anode. In such conditions mid-gap streamers are readily observed (Dunnington, 1931; White, 1934; Raether, 1937, 39), and are accompanied by a reduction in the time of formation of the spark in accordance with expectation from theory.

Confirmation of the basic ideas of the new theory is found partly in Raether's observations on the development of avalanches and sparks in a cloud chamber (1939). His photographs show firstly that the transition from avalanche to streamer develops about ten times as fast. They also show that the average speed of progress of an avalanche is about 2.10^7 cm.sec^{-1}, in accordance with the speed estimated in the derivation of expression (3), whilst the return positive streamer develops about ten times as fast. It follows that the time of formation of a spark in a uniform field gap at the minimum breakdown potential is then governed by the transit time of an electron across the gap.

In order that the breakdown of a uniform gap may occur at the minimum sparking potential, the electron avalanche which leads to the formation of a positive streamer must be initiated by an electron leaving the cathode and the avalanche must travel the full length of the gap and attain an extremely large ion density. Only one such electron avalanche is required, but owing to the random

character of the ionization processes, several avalanches may have to be initiated before one is successful in causing a spark. With a potential slightly in excess of the minimum to cause breakdown, the enhanced ionization in the avalanches makes it more probable that any single avalanche will lead to breakdown of the gap, and Strigel (1932) has shown that the statistical time-lag of a gap corresponds very closely to the average time interval between the emission of individual electrons from the cathode. In order to obtain reproducible conditions in the practically important case when the application of the electrode potential is limited to 1 μ sec. or less, the emission from the cathode of at least one primary electron must occur during this interval. The usual methods by which such primary electrons are created include irradiation of the gap by radioactive materials (van Cauwenberghe, 1937; Dillon, 1939; Berkey, 1934, 40) or else by the ultra-violet light emitted by a mercury-arc lamp (Rogowski and Tamm, 1928; Nord, 1935; Loeb, 1939). By the latter method, a photoelectric current of 10^{-12} amp. per cm^2 of cathode can readily be obtained in discharge-gaps of the type normally used in the laboratory. This corresponds to the liberation of one electron per $0{\cdot}16\,\mu$sec. per cm^2 of cathode, and erratic results may be expected in a gap where the cathode area is much less than 1 cm^2 when impulse voltages of duration 1 μ sec. only are applied. Thus it is not surprising that irregular breakdown is obtained in the measurement of impulse voltages using small electrodes (Strigel, 1932, 38), even when irradiated by the light from a mercury-arc lamp. The requirement for reproducibility is then a more intense irradiation of the gap: the radiation from an adjacent spark appears to be satisfactory (Wynn-Williams, 1926; Street and Beams, 1931; White, 1936; Raether, 1938; Garfitt, 1942; Meek, 1942).

An alternative method by which a high rate of electron liberation from the cathode may be produced has recently been discussed by Slepian and Berkey (1940), who investigated the effect of various small, insulating, sharp-edged particles when placed on the cathode. It was found that for particles of average linear dimensions 10^{-2} to 10^{-3} cm., the statistical time-lag of the gap was materially reduced, due probably to the increased initiatory electron current from the cathode caused by the localized field distortions. A similar effect has been noted by Paetow (1939), who observed that for cathodes on which insulating oxides were present, a single discharge produced conditions which led to a considerable reduction in the statistical time-lag for subsequent discharges. This would appear to be due to the charge developed on the particles of non-conducting material on the cathode by positive-ion bombardment or by photo-electric effect, and Paetow has postulated that field emission may then take place, though it is also possible that the particles may themselves break down and the resultant discharge furnish the electrons required in the main gap (Loeb, 1939).

When a gap is subjected to the intense ultra-violet radiations emitted by a nearby spark, an actual lowering of the breakdown potential of the gap is observed if the light source is brought sufficiently close to the gap under observation

(White, 1935; Rogowski and Wallraff, 1935; Brinkman, 1939, 40). This lowering may be explained by the fact that the large photo-electric current gives rise to a positive-ion space-charge distributed across the gap, which builds up even though the potential across the gap is less than that required to cause spark-over under normal conditions (Franck and von Hippel, 1929; Varney, White, Loeb and Posin, 1935; Rogowski and Wallraff, 1937; Seeliger, 1939; Rogowski, 1939; Meek, 1940; Loeb and Meek, 1941). Consequently this positive space-charge may distort the applied field so that $\int_0^d \alpha_b dz$ is increased, and, when this integral replaces $\alpha_b d$ in (4), this equation is satisfied by a lower value of X. Seeliger (1939) has shown that $\int_0^d \alpha dz$ for air may be increased in this manner provided that X/p is less than about 300.

The case of non-uniform gap fields presents difficulties similar to those mentioned in § 3 in relation to the estimation of the ionization attributable to electron impact in the cathode dark space. It is then necessary to substitute $\int_0^d \alpha_b dz$ for the term $\alpha_b d$ in (4). But the subsequent substitution of $\int_0^d \alpha dz$ for $\int_0^d \alpha_b dz$ leads to the additional question as to whether the data for α which relate to a uniform field are applicable to non-uniform fields; there seems to be no conclusive answer to this problem despite the various attempts made (see § 3). But again, despite the uncertainties undoubtedly involved, the use of $\int_0^d \alpha dz$ in (4) to determine the breakdown potential for sphere-gaps and other electrode arrangements leads to remarkable agreement between theory and experiment.

For those gaps in which corona discharge is observed at a potential lower than that required to cause breakdown, only the corona onset potential can be calculated from (4) (Loeb and Meek, 1941). The difference between corona on positive and negative conductors is also accounted for: in the former case, the discharge is initiated by an avalanche which arrives at the anode with a sufficient ion density to produce a corona streamer, whereas for corona at a negative electrode, the electron avalanche proceeds outward from the electrode into a diminishing electric field and gives rise to a " mid-gap streamer " which progresses back to the cathode and also towards the anode if conditions are propitious.

To summarize, it would appear that while the conditions which govern the initiation of streamers now seem to be understood, no complete explanation of the propagation of streamers has yet been forthcoming except in a qualitative manner (Loeb and Kip, 1939; Loeb, 1939). Once a streamer has formed in a uniform field, the consequent distortion of the already high field existing in the gap is sufficient to enable the propagation of the streamer across the gap. However, in long non-uniform gaps, such as the point-plane arrangement. or

that which obtains in the case of lightning discharges, the gradient in which the streamer (or leader stroke) develops is initially only some few hundred volts per cm. over a considerable proportion of the gap. Bruce (1941) has given an explanation of the development of lightning leader strokes, in low external fields, in terms of observations pertaining to the glow-to-arc transition between metal electrodes, but the validity of this explanation is questionable because such a transition seems to be governed by cathode mechanisms, while the leader stroke is a gas-dependent phenomenon. However, it appears that as the leader stroke advances, it must carry forward a localized intense field about its tip, in order that the advance may be maintained. This requires an excess charge of the appropriate sign in the channel, and the gradient along the channel should then be such that the electron drift speed is sufficient to ensure the continuous creation of this excess charge as the leader advances (Meek, 1941).

GENERAL REFERENCES

Druyvesteyn, M. J. and Penning, F. M., 1940. *Rev. Mod. Phys.* **12**, 87.

von Engel, A. and Steenbeck, M., 1932. *Elektrische Gasentladungen*, **1** (Berlin: J. Springer).

von Engel, A. and Steenbeck, M., 1934. *Elektrische Gasentladungen* **2** (Berlin: J. Springer).

Loeb, L. B., 1939. *Fundamental Processes of Electrical Discharges in Gases* (New York: John Wiley and Sons).

Thomson, J. J. and Thomson, G. P., 1933. *Conduction of Electricity through Gases* (Cambridge University Press).

Townsend, J. S. E., 1915. *Electricity in Gases* (Oxford University Press).

OTHER REFERENCES

§2. *The statistical theory of discharge reactions* (*R. W. L.*)

Bailey, V. A. and Higgs, A. J., 1929. *Phil. Mag.* **7** (vii), 277.

Bradbury, N. E., 1933. *Phys. Rev.* **44**, 833.

Brode, R. B., 1933. *Rev. Mod. Phys.* **5**, 257.

Druyvesteyn, M. J., 1933. *Z. Phys.* **81**, 571.

Druyvesteyn, M. J., 1936. *Physica*, **3**, 65.

Emeléus, K. G. and Lunt, R. W., 1936. *Trans. Faraday Soc.* **32**, 1504.

Emeléus, K. G., Lunt, R. W. and Meek, C. A., 1936. *Proc. Roy. Soc.* A, **156**, 394.

Francis, V. J. and Jenkins, H. G., 1940. *Rep. Prog. Phys.* **7**, 230.

Harris, W. A., 1934. *Phil. Mag.* **17**, 131.

Killian, T. J., 1930. *Phys. Rev.* **35**, 1238.

Kirkby, Rev. P. J., 1902. *Phil. Mag.* **3** (vi), 212.

Kirkby, Rev. P. J., 1911. *Proc. Roy. Soc.* A, **85**, 151.

Lunt, R. W., and Meek, C. A., 1936. *Proc. Roy. Soc.* A, **157**, 146.

Lunt, R. W., Meek, C. A. and Smith, E. C. W., 1937. *Proc. Roy. Soc.* A, **158**, 729.

Lunt, R. W. and Swindell, G. E., 1940. *Trans. Faraday Soc.* **36**, 1072.

Lunt, R. W., Pearson, T. G., and Topley, B., 1940. *Trans. Faraday Soc.* **36**, 1087.

Tate, J. T. and Smith, P. T., 1932. *Phys. Rev.* **27**, 724.

Tonks, L., 1936. Private communication.

Townsend, J. S. E., 1901. *Phil. Mag.* **1** (vi), 198 and 630.

Townsend, J. S. E., 1903. *Phil. Mag.* **5** (vi), 389.

Townsend, J. S. E., 1925. *Motion of Electrons in Gases* (Oxford University Press).

§ 3. *The mechanism of the cathode fall in glow discharges* (*A. v. E.*)

ARNOT, F. L., 1937. *Proc. Roy. Soc.* A, **158**, 137.
ARNOT, F. L. and MILLIGAN, J. C., 1936. *Proc. Roy. Soc.* A, **156**, 538.
ASTON, F. W., 1911. *Proc. Roy. Soc.* A, **84**, 526.
BÄR, R., 1927. *Handb. d. Physik*, **14**, 203.
BREWER, A. K. and WESTHAVER, J. W., 1937. *J. Appl. Phys.* **8**, 779.
BROSE, E., 1919. *Ann. Phys., Lpz.*, **58**, 731.
CAMPAN, T. J., 1934. *Z. Phys.* **91**, 111.
CHIPLONKAR, V. T., 1940. *Proc. Indian Acad. Sci., Bangalore*, A, **12**, 440.
COMPTON, K. T. and LANGMUIR, I., 1930. *Rev. Mod. Phys.* **2**, 221.
COMPTON, K. T. and MORSE, M., 1928. *Phys. Rev.* **30**, 305.
DARROW, K. K., 1932. *Electrical Phenomena in Gases* (Baltimore : Williams and Wilkins).
DAUVVESTEYN, M. J., 1930. *Z. Phys.* **62**, 764.
DRUYVESTEYN, M. J., 1938. *Physica*, **5**, 875.
DRUYVESTEYN, M. J. and WARMOLTZ, N., 1931. *Z. Phys.* **68**, 378.
EMELÉUS, K. G. and HARRIS, N. L., 1927. *Phil. Mag.* **4**, 49.
EMELÉUS, K. G., BROWN, W. L. and COWAN, H. McN., 1934. *Phil. Mag.* **17**, 146.
EMELÉUS, K. G. and DUFFENDACK, O. S., 1935. *Phys. Rev.* **47**, 460.
EMELÉUS, K. G. and SAYERS, J., 1938. *Proc. R. Irish Acad.* **44**, 87.
VON ENGEL, A., SEELIGER, R. and STEENBECK, M., 1933. *Z. Phys.* **85**, 144.
ERNST, H. W., 1935. *Helv. Phys. Acta*, **8**, 381.
FISCHER, H., 1936. *Ann. Phys., Lpz.*, **27**, 81.
FISCHER, H., 1938. *Z. Phys.* **110**, 197.
FRANCIS, V. J. and JENKINS, K. G., 1940. *Rep. Prog. Phys.* **7**, 230.
FREY, B., 1932. *See* DARROW, K. K., 1932, p. 420.
FUCKS, W., 1933. *Z. Phys.* **87**, 139.
GEHRTS, A., 1936. *Z. tech. Phys.* **17**, 393.
GEIGER, K., 1937. *Z. Phys.* **106**, 17.
GERTHSEN, H., 1928. *Ann. Phys., Lpz.*, **85**, 881.
GÜNTHERSCHULZE, A., 1928. *Z. Phys.* **49**, 358.
GÜNTHERSCHULZE, A., 1930 a. *Z. Phys.* **59**, 433.
GÜNTHERSCHULZE, A., 1930 b. *Z. Phys.* **62**, 600.
GÜNTHERSCHULZE, A. and KELLER, F., 1931. *Z. Phys.* **68**, 162.
KALLMANN, H. and ROSEN, B., 1930. *Z. Phys.* **64**, 806.
KENTY, C., 1933. *Phys. Rev.* **44**, 891.
KURZKE, H., 1936. *Ann. Phys., Lpz.*, **25**, 688.
LANGMUIR, I. and TONKS, L., 1929. *Phys. Rev.* **33**, 195 and 990.
MASSEY, H. S. W., 1938. *Negative Ions* (Cambridge University Press).
MOTT, N. F. and MASSEY, H. S. W., 1933. *Theory of Atomic Collisions* (Oxford : Clarendon Press).
OLIPHANT, M. L. E., 1929. *Proc. Roy. Soc.* A, **124**, 228.
OLIPHANT, M. L. E., 1930. *Proc. Roy. Soc.* A, **127**, 373.
PAETOW, H. and WALCHE, W., 1938. *Z. Phys.* **110**, 69.
PENNING, M., 1932. *Phys. Z.* **33**, 816.
PIKE, E. W., 1934. *Z. Phys.* **90**, 319.
ROGOWSKI, W., 1932. *Arch. Elektrotech.* **26** 643.
ROGOWSKI, W., 1939. *Z. Phys.* **114**, 1.
RUNGE, I., 1930. *Z. Phys.* **61**, 174.
SCHADE, R., 1938. *Z. Phys.* **108**, 353.
SCHERZER, O., 1939. *Arch. Elektrotech.* **33**, 207.
SEELIGER, R., 1934. *Physik d. Gasentladungen*, 2nd ed. (Leipzig : Barth).
SEELIGER, R., 1932. *Phys. Z.* **33**, 273.
SEELIGER, R., 1940. *Z. Phys.* **115**, 111.

Sloane, R. H. and Emeléus, K. G., 1931. *Phil. Mag.* **12**, 73.
Smith, R. A., 1938. *Proc. Roy. Soc.* A, **168**, 19.
Spiwak, E. and Reichrudel, G., 1936. *Phys. Z. Sowjet*, **10**, 121.
Steubing, N., 1932. *Ann. Phys., Lpz.*, **10**, 296.
Timoshenko, G., 1941. *J. Appl. Phys.* **12**, 69.
Townsend, J. S., 1939. *Phil. Mag.* **28**, 111.
Tyndall, A. M., 1938. *The Mobility of Positive Ions in Gases* (Cambridge University Press).
van Voorhis, C. C. and Shenstone, A. G., 1941. *Rev. Sci. Instrum.* **12**, 257.
Wehnelt, A. and Kurzke, H., 1934. *Ann. Phys., Lpz.*, **20**, 919.
Weizel, W. and Fischer, H., 1935. *Ann. Phys., Lpz.*, **24**, 209.
Weizel, W., Rompe, R. and Schön, M., 1939 a. *Z. Phys.* **112**, 339.
Weizel, W., Rompe, R. and Schön, M., 1939 b. *Z. Phys.* **113**, 87 and 730.
Wolf, F., 1937. *Ann. Phys., Lpz.*, **30**, 313.

§4. *Recent developments in the theory of the spark discharge* (*J. M. M.*)

Berkey, W. E., 1934. *Elect. J.* **31**, 101.
Berkey, W. E., 1940. *Trans. Amer. Inst. Elect. Engrs.* **59**, 429.
Brinkman, C., 1939 a. *Z. Phys.* **111**, 737.
Brinkman, C., 1939 b. *Arch. Elektrotech.* **33**, 1, 121.
Bruce, C. E. R., 1941. *Nature, Lond.*, **147**, 805.
van Cauwenberghe, R., 1937. *Bull. Soc. franç. Elect.*, October.
Dillon, J. H., 1940. *J. Appl. Phys.* **11**, 291.
Dunnington, F. G., 1931. *Phys. Rev.* **38**, 1535.
Franck, J. and von Hippel, A., 1929. *Z. Phys.* **57**, 695.
Garfitt, D. E. M., 1942. *Proc. Phys. Soc.* In press.
Holm, R., 1932. *Z. Phys.* **75**, 171.
Holst, G. and Oosterhuis, E., 1923. *Phil. Mag.* **46**, 1117.
Kapzov, N., 1932. *Z. Phys.* **75**, 380.
Loeb, L. B., 1928. *J. Franklin Inst.* **205**, 305.
Loeb, L. B., 1936. *Rev. Mod. Phys.* **8**, 267.
Loeb, L. B. and Kip, A. F., 1939. *J. Appl. Phys.* **10**, 142.
Loeb, L. B. and Meek, J. M., 1940. *J. Appl. Phys.* **11**, 438, 459.
Loeb, L. B. and Meek, J. M., 1941. *The Mechanism of the Electric Spark* (Oxford University Press).
Meek, J. M., 1940 a. *Phys. Rev.* **57**, 722.
Meek, J. M., 1940 b. *J. Franklin Inst.* **230**, 229.
Meek, J. M., 1940 c. *Proc. Phys. Soc.* **52**, 547, 822.
Meek, J. M., 1941. *Nature, Lond.*, **148**, 437.
Meek, J. M., 1942. *J. Inst. Elect. Engrs.* In press.
Morgan, J. D., 1927. *Phil. Mag.* **4**, 91.
Mott, N. F. and Massey, H. S. W., 1933. *Theory of Atomic Collisions* (Oxford : Clarendon Press).
Nord, G., 1935. *Trans. Amer. Inst. Elect. Engrs.* **54**, 955.
Paetow, H., 1939. *Z. Phys.* **111**, 770.
Raether, H., 1937. *Z. Phys.* **107**, 91.
Raether, H., 1938. *Z. Phys.* **110**, 611.
Raether, H., 1939. *Z. Phys.* **112**, 464.
Rogowski, W., 1928. *Arch. Elektrotech.* **20**, 99.
Rogowski, W., 1939. *Z. Phys.* **114**, 1.
Rogowski, W. and Tamm, R., 1928. *Arch. Elektrotech.* **20**, 107.
Rogowski, W. and Wallraff, A., 1935. *Z. Phys.* **97**, 758.
Rogowski, W. and Wallraff, A., 1938. *Z. Phys.* **108**, 1.

Sämmer, J. J., 1933. *Z. Phys.* **83**, 11, 814.
Schade, R., 1939. *Z. Phys.* **111**, 437.
Schumann, W. O., 1930. *Z. tech. Phys.* **11**, 194.
Seeliger, R., 1934. *Physik d. Gasentladungen* (Leipzig : Barth).
Skilling, H. H. and Brenner, W. O., 1941. *Trans. Amer. Inst. Elect. Engrs.* **60**, 132.
Slepian, J. and Berkey, W. E., 1940. *J. Appl. Phys.* **11**, 765.
Street, J. C. and Beams, J. W., 1931. *Phys. Rev.* **38**, 416.
Strigel, R., 1932. *Wiss. Veröff. Siemens-Konz.* **11**, 52.
Strigel, R., 1938. *Elektrotech. Z.* **59**, 1, 33, 60.
Taylor, J., 1927. *Proc. Roy. Soc.* A, **114**, 73.
Trump, J. G., Safford, F. J. and Cloud, R. W., 1941. *Trans. Amer. Inst. Elect. Engrs.* **60**, 132.
Varney, R. N., Loeb, L. B. and Haseltine, W. R., 1940. *Phil. Mag.* **29**, 379.
Varney, R. N., White, H. J., Loeb, L. B. and Posin, D. Q., 1935. *Phys. Rev.* **48**, 878.
White, H. J., 1934. *Phys. Rev.* **46**, 99.
White, H. J., 1935. *Phys. Rev.* **48**, 113.
White, H. J., 1936. *Phys. Rev.* **49**, 507.
Wynn-Williams, C. E., 1926. *Phil. Mag.* **1**, 353.